

DEX-0214

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

EF

(19) World Intellectual Property Organization  
International Bureau(43) International Publication Date  
31 January 2002 (31.01.2002)

PCT

(10) International Publication Number  
WO 02/08278 A2(51) International Patent Classification<sup>7</sup>: C07K 14/47

Clara, CA 95050 (US). CHEN, Seiyu [US]; 160 Mira Street, Foster City, CA 94404 (US).

(21) International Application Number: PCT/US01/22949

(74) Agents: LICATA, Jane, Massey et al.; Licata &amp; Tyrrell P.C., 66 E. Main Street, Marlton, NJ 08053 (US).

(22) International Filing Date: 20 July 2001 (20.07.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/219,834 21 July 2000 (21.07.2000) US(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:  
US 60/219,834 (CIP)  
Filed on 21 July 2000 (21.07.2000)(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).(71) Applicant (*for all designated States except US*): DI-ADEXUS, INC. [US/US]; 343 Oyster Point Boulevard, South San Francisco, CA 94080 (US).**Published:**

— without international search report and to be republished upon receipt of that report

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): MACINA, Roberto, A. [AR/US]; 4118 Crescendo Avenue, San Jose, CA 95136 (US). NAIR, Manoj [US/US]; 2491 Ramke Place, Santa

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.



WO 02/08278 A2

(54) Title: COMPOSITIONS AND METHODS RELATING TO LUNG SPECIFIC GENES

(57) Abstract: The invention relates to LSG polypeptides, polynucleotides encoding the polypeptides, methods for producing the polypeptides, in particular by expressing the polynucleotides, and agonists and antagonists of the polypeptides. The invention further relates to methods for utilizing such polynucleotides, polypeptides, agonists and antagonists for applications, which relate, in part, to research, diagnostic and clinical arts.

DEX-0514

Et

COMPOSITIONS AND METHODS  
RELATING TO LUNG SPECIFIC GENES

INTRODUCTION

This application claims the benefit of priority from  
5 U.S. Provisional Application Serial No. 60/219,834, filed  
July 21, 2000, which is herein incorporated in its  
entirety.

FIELD OF THE INVENTION

The present invention relates to newly identified  
10 nucleic acids and polypeptides present in normal and  
neoplastic lung cells, including fragments, variants and  
derivatives of the nucleic acids and polypeptides. The  
present invention also relates to antibodies to the  
polypeptides of the invention, as well as agonists and  
15 antagonists of the polypeptides of the invention. The  
invention also relates to compositions comprising the  
nucleic acids, polypeptides, antibodies, variants,  
derivatives, agonists and antagonists of the invention and  
methods for the use of these compositions. These uses  
20 include identifying, diagnosing, monitoring, staging,  
imaging and treating lung cancer and non-cancerous disease  
states in lung, identifying lung tissue, monitoring and  
modifying lung embryonic development and differentiation,  
and identifying and/or designing agonists and antagonists  
25 of polypeptides of the invention. The uses also include  
gene therapy, production of transgenic animals and cells,  
and production of engineered lung tissue for treatment and  
research.

- 2 -

## BACKGROUND OF THE INVENTION

Throughout the last hundred years, the incidence of lung cancer has steadily increased, so much so that now in many countries, it is the most common cancers. In fact, lung cancer is the second most prevalent type of cancer for both men and women in the United States and is the most common cause of cancer death in both sexes. Lung cancer deaths have increased ten-fold in both men and women since 1930, primarily due to an increase in cigarette smoking, but also due to an increased exposure to arsenic, asbestos, chromates, chloromethyl ethers, nickel, polycyclic aromatic hydrocarbons and other agents. See Scott, Lung Cancer: A Guide to Diagnosis and Treatment, Addicus Books (2000) and Alberg et al., in Kane et al. (eds.) Biology of Lung Cancer, pp. 11-52, Marcel Dekker, Inc. (1998). Lung cancer may result from a primary tumor originating in the lung or a secondary tumor which has spread from another organ such as the bowel or breast. Although there are over a dozen types of lung cancer, over 90% fall into two categories: small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC). See Scott, *supra*. About 20-25% of all lung cancers are characterized as SCLC, while 70-80% are diagnosed as NSCLC. *Id.* A rare type of lung cancer is mesothelioma, which is generally caused by exposure to asbestos, and which affects the pleura of the lung. Lung cancer is usually diagnosed or screened for by chest x-ray, CAT scans, PET scans, or by sputum cytology. A diagnosis of lung cancer is usually confirmed by biopsy of the tissue. *Id.*

SCLC tumors are highly metastatic and grow quickly. By the time a patient has been diagnosed with SCLC, the cancer has usually already spread to other parts of the body, including lymph nodes, adrenals, liver, bone, brain and bone marrow. See Scott, *supra*; Van Houtte et al. (eds.), Progress and Perspective in the Treatment of Lung



- 3 -

Cancer, Springer-Verlag (1999). Because the disease has usually spread to such an extent that surgery is not an option, the current treatment of choice is chemotherapy plus chest irradiation. See Van Houtte, *supra*. The stage  
5 of disease is a principal predictor of long-term survival. Less than 5% of patients with extensive disease that has spread beyond one lung and surrounding lymph nodes, live longer than two years. *Id.* However, the probability of five-year survival is three to four times higher if the  
10 disease is diagnosed and treated when it is still in a limited stage, i.e., not having spread beyond one lung. *Id.*

NSCLC is generally divided into three types: squamous cell carcinoma, adenocarcinoma and large cell  
15 carcinoma. Both squamous cell cancer and adenocarcinoma develop from the cells that line the airways; however, adenocarcinoma develops from the goblet cells that produce mucus. Large cell lung cancer has been thus named because the cells look large and rounded when viewed  
20 microscopically, and generally are considered relatively undifferentiated. See Yesner, *Atlas of Lung Cancer*, Lippincott-Raven (1998).

Secondary lung cancer is a cancer initiated elsewhere in the body that has spread to the lungs. Cancers that  
25 metastasize to the lung include, but are not limited to, breast cancer, melanoma, colon cancer and Hodgkin's lymphoma. Treatment for secondary lung cancer may depend upon the source of the original cancer. In other words, a lung cancer that originated from breast cancer may be more  
30 responsive to breast cancer treatments and a lung cancer that originated from the colon cancer may be more responsive to colon cancer treatments.

The stage of a cancer indicates how far it has spread and is an important indicator of the prognosis. In  
35 addition, staging is important because treatment is often

- 4 -

decided according to the stage of a cancer. SCLC is divided into two stages: limited disease, i.e., cancer that can only be seen in one lung and in nearby lymph nodes; and extensive disease, i.e., cancer that has spread  
5 outside the lung to the chest or to other parts of the body. For most patients with SCLC, the disease has already progressed to lymph nodes or elsewhere in the body at the time of diagnosis. See Scott, *supra*. Even if spreading is not apparent on the scans, it is likely that some cancer  
10 cells may have spread away and traveled through the bloodstream or lymph system. In general, chemotherapy with or without radiotherapy is often the preferred treatment. The initial scans and tests done at first will be used later to see how well a patient is responding to treatment.

15 In contrast, non-small cell cancer may be divided into four stages. Stage I is highly localized cancer with no cancer in the lymph nodes. Stage II cancer has spread to the lymph nodes at the top of the affected lung. Stage III cancer has spread near to where the cancer started.  
20 This can be to the chest wall, the covering of the lung (pleura), the middle of the chest (mediastinum) or other lymph nodes. Stage IV cancer has spread to another part of the body. Stage I-III cancer is usually treated with surgery, with or without chemotherapy. Stage IV cancer is  
25 usually treated with chemotherapy and/or palliative care.

A number of chromosomal and genetic abnormalities have been observed in lung cancer. In NSCLC, chromosomal aberrations have been described on 3p, 9p, 11p, 15p and 17p, and chromosomal deletions have been seen on  
30 chromosomes 7, 11, 13 and 19. See Skarin (ed.), *Multimodality Treatment of Lung Cancer*, Marcel Dekker, Inc. (2000); Gemmill et al., pp. 465-502, in Kane, *supra*; Bailey-Wilson et al., pp. 53-98, in Kane, *supra*. Chromosomal abnormalities have been described on 1p, 3p,  
35 5q, 6q, 8q, 13q and 17p in SCLC. *Id.* In addition, the

- 5 -

loss of the short arm of chromosome 3p has also been seen in greater than 90% of SCLC tumors and approximately 50% of NSCLC tumors. *Id.*

A number of oncogenes and tumor suppressor genes have been implicated in lung cancer. See Mabry, pp. 391-412, in Kane, *supra* and Sclafani et al., pp. 295-316, in Kane, *supra*. In both SCLC and NSCLC, the p53 tumor suppressor gene is mutated in over 50% of lung cancers. See Yesner, *supra*. Another tumor suppressor gene, FHIT, which is found on chromosome 3p, is mutated by tobacco smoke. *Id.*; Skarin, *supra*. In addition, more than 95% of SCLCs and approximately 20-60% of NSCLCs have an absent or abnormal retinoblastoma (Rb) protein, another tumor suppressor gene. The ras oncogene (particularly K-ras) is mutated in 20-30% of NSCLC specimens and the c-erbB2 oncogene is expressed in 18% of stage 2 NSCLC and 60% of stage 4 NSCLC specimens. See Van Houtte, *supra*. Other tumor suppressor genes that are found in a region of chromosome 9, specifically in the region of 9p21, are deleted in many cancer cells, including p16<sup>INK4A</sup> and p15<sup>INK4B</sup>. See Bailey-Wilson, *supra*; Sclafani et al., *supra*. These tumor suppressor genes may also be implicated in lung cancer pathogenesis.

In addition, many lung cancer cells produce growth factors that may act in an autocrine fashion on lung cancer cells. See Siegfried et al., pp. 317-336, in Kane, *supra*; Moody, pp. 337-370, in Kane, *supra* and Heasley et al., 371-390, in Kane, *supra*. In SCLC, many tumor cells produce gastrin-releasing peptide (GRP), which is a proliferative growth factor for these cells. See Skarin, *supra*. Many NSCLC tumors express epidermal growth factor (EGF) receptors, allowing NSCLC cells to proliferate in response to EGF. Insulin-like growth factor (IGF-I) is elevated in greater than 95% of SCLC and greater than 80% of NSCLC tumors; it is thought to function as an autocrine growth

- 6 -

factor. *Id.* Finally, stem cell factor (SCF, also known as steel factor or kit ligand) and c-Kit (a proto-oncoprotein tyrosine kinase receptor for SCF) are both expressed at high levels in SCLC, and thus may form an autocrine loop  
5 that increases proliferation. *Id.*

Although the majority of lung cancer cases are attributable to cigarette smoking, most smokers do not develop lung cancer. Epidemiological evidence has suggested that susceptibility to lung cancer may be  
10 inherited in a Mendelian fashion, and thus have an inherited genetic component. Bailey-Wilson, *supra*. Thus, it is thought that certain allelic variants at some genetic loci may affect susceptibility to lung cancer. *Id.* One way to identify which allelic variants are likely to be  
15 involved in lung cancer susceptibility, as well as susceptibility to other diseases, is to look at allelic variants of genes that are highly expressed in lung.

The lung is also susceptible to a number of other debilitating diseases, including, without limitation,  
20 emphysema, pneumonia, cystic fibrosis and asthma. See Stockley (ed.), *Molecular Biology of the Lung, Volume I: Emphysema and Infection*, Birkhauser Verlag (1999), hereafter Stockley I, and Stockley (ed.), *Molecular Biology of the Lung, Volume II: Asthma and Cancer*, Birkhauser  
25 Verlag (1999), hereafter Stockley II. The cause of many these disorders is still not well understood and there are few, if any, good treatment options for many of these noncancerous lung disorders. Thus, there remains a need to understand various noncancerous lung disorders and to  
30 identify treatments for these diseases.

In yet another aspect, the development and differentiation of the lung tissue is important during embryonic development. All of the epithelial cells of the respiratory tract, including those of the lung and bronchi,  
35 are derived from the primitive endodermal cells that line

- 7 -

the embryonic outpouching. See Yesner, *supra*. During embryonic development, multipotent endodermal stem cells differentiate into many different types of specialized cells, which include ciliated cells for moving inhaled particles, goblet cells for producing mucus, Kulchitsky's cells for endocrine function, and Clara cells and type II pneumocytes for secreting surfactant protein. *Id.* Improper development and differentiation may cause respiratory disorders and distress in infants, particularly in premature infants, whose lungs cannot produce sufficient surfactant when they are born. Further, some lung cancer cells, particularly small cell carcinomas, appear multipotent, and can spontaneously differentiate into a number of cell types, including small cell carcinoma, adenocarcinoma and squamous cell carcinoma. *Id.* Thus, a better understanding of lung development and differentiation may help facilitate understanding of lung cancer initiation and progression.

Accordingly, there is a great need for more sensitive and accurate methods for predicting whether a person is likely to develop lung cancer, for diagnosing lung cancer, for monitoring the progression of the disease, for staging the lung cancer, for determining whether the lung cancer has metastasized and for imaging the lung cancer. There is also a need for better treatment of lung cancer. Further, there is also a great need for diagnosing and treating noncancerous lung disorders such as emphysema, pneumonia, lung infection, pulmonary fibrosis, cystic fibrosis and asthma. There is also a need for compositions and methods of using them that can be used to identify lung tissue for forensic purposes and for determining whether a particular cell or tissue exhibits lung-specific characteristics.

In the present invention, methods are provided for detecting, diagnosing, monitoring, staging, prognosticating, imaging and treating lung cancer via lung

- 8 -

specific genes referred to herein as LSGs. For purposes of the present invention, LSG refers, among other things, to native protein expressed by the gene comprising a polynucleotide sequence of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, or 22 or a contig of SEQ ID NO: 19 or 21 as depicted in SEQ ID NO: 37, or 38, respectively. By "LSG" it is also meant herein polynucleotides which, due to degeneracy in genetic coding, comprise variations in nucleotide sequence as compared to SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 37, 38, 39 or 40 but which still encode the same polypeptide. Exemplary amino acid sequences for LSG polypeptides are set forth in SEQ ID NO: 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55 and 56. In the alternative, what is meant by LSG as used herein, means the native mRNA encoded by the gene comprising the polynucleotide sequence of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21 or 22, levels of the gene comprising the polynucleotide sequence of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, or 22 or levels of a polynucleotide which is capable of hybridizing under stringent conditions to the antisense sequence of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 37, or 38.

Other objects, features, advantages and aspects of the present invention will become apparent to those of skill in the art from the following description. It should be understood, however, that the following description and the specific examples, while indicating preferred embodiments of the invention are given by way of illustration only. Various changes and modifications within the spirit and scope of the disclosed invention will become readily apparent to those skilled in the art from

- 9 -

reading the following description and from reading the other parts of the present disclosure.

#### SUMMARY OF THE INVENTION

Toward these ends, and others, it is an object of the present invention to provide LSGs comprising a polynucleotide of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 37 or 38 a protein expressed by a polynucleotide of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 37 or 38 or a variant thereof which expresses the protein; or a polynucleotide which is capable of hybridizing under stringent conditions to the antisense sequence of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 37 or 38. Exemplary LSG polypeptides of the present invention are depicted in SEQ ID NO: 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55 or 56.

It is another object of the present invention to provide a method for diagnosing the presence of lung cancer by analyzing for changes in levels of LSG in cells, tissues or bodily fluids compared with levels of LSG in preferably the same cells, tissues, or bodily fluid type of a normal human control, wherein a change in levels of LSG in the patient versus the normal human control is associated with lung cancer.

Further provided is a method of diagnosing metastatic lung cancer in a patient having lung cancer which is not known to have metastasized by identifying a human patient suspected of having lung cancer that has metastasized; analyzing a sample of cells, tissues, or bodily fluid from such patient for LSG; comparing the LSG levels in such cells, tissues, or bodily fluid with levels of LSG in preferably the same cells, tissues, or bodily fluid type of a normal human control, wherein an increase in LSG levels

- 10 -

in the patient versus the normal human control is associated with lung cancer which has metastasized.

Also provided by the invention is a method of staging lung cancer in a human which has such cancer by identifying  
5 a human patient having such cancer; analyzing a sample of cells, tissues, or bodily fluid from such patient for LSG; comparing LSG levels in such cells, tissues, or bodily fluid with levels of LSG in preferably the same cells, tissues, or bodily fluid type of a normal human control  
10 sample, wherein an increase in LSG levels in the patient versus the normal human control is associated with a cancer which is progressing and a decrease in the levels of LSG is associated with a cancer which is regressing or in remission.

15 Further provided is a method of monitoring lung cancer in a human having such cancer for the onset of metastasis. The method comprises identifying a human patient having such cancer that is not known to have metastasized; periodically analyzing a sample of cells,  
20 tissues, or bodily fluid from such patient for LSG; comparing the LSG levels in such cells, tissue, or bodily fluid with levels of LSG in preferably the same cells, tissues, or bodily fluid type of a normal human control sample, wherein an increase in LSG levels in the patient  
25 versus the normal human control is associated with a cancer which has metastasized.

Further provided is a method of monitoring the change in stage of lung cancer in a human having such cancer by looking at levels of LSG in a human having such cancer.  
30 The method comprises identifying a human patient having such cancer; periodically analyzing a sample of cells, tissues, or bodily fluid from such patient for LSG; comparing the LSG levels in such cells, tissue, or bodily fluid with levels of LSG in preferably the same cells,  
35 tissues, or bodily fluid type of a normal human control



- 11 -

sample, wherein an increase in LSG levels in the patient versus the normal human control is associated with a cancer which is progressing and a decrease in the levels of LSG is associated with a cancer which is regressing or in  
5 remission.

Further provided are methods of designing new therapeutic agents targeted to a LSG for use in imaging and treating lung cancer. For example, in one embodiment, therapeutic agents such as antibodies targeted against LSG  
10 or fragments of such antibodies can be used to treat, detect or image localization of LSG in a patient for the purpose of detecting or diagnosing a disease or condition. In this embodiment, an increase in the amount of labeled antibody detected as compared to normal tissue would be  
15 indicative of tumor metastases or growth. Such antibodies can be polyclonal, monoclonal, or omniclonal or prepared by molecular biology techniques. The term "antibody", as used herein and throughout the instant specification is also meant to include aptamers and single-stranded  
20 oligonucleotides such as those derived from an *in vitro* evolution protocol referred to as SELEX and well known to those skilled in the art. Antibodies can be labeled with a variety of detectable and therapeutic labels including, but not limited to, radioisotopes and paramagnetic metals.  
25 Therapeutic agents such as small molecules and antibodies which decrease the concentration and/or activity of LSG can also be used in the treatment of diseases characterized by overexpression of LSG. Such agents can be readily identified in accordance with teachings herein.  
30 Other objects, features, advantages and aspects of the present invention will become apparent to those of skill in the art from the following description. It should be understood, however, that the following description and the specific examples, while indicating preferred  
35 embodiments of the invention, are given by way of

- 12 -

illustration only. Various changes and modifications within the spirit and scope of the disclosed invention will become readily apparent to those skilled in the art from reading the following description and from reading the  
5 other parts of the present disclosure.

#### GLOSSARY

The following illustrative explanations are provided to facilitate understanding of certain terms used frequently herein, particularly in the examples. The  
10 explanations are provided as a convenience and are not limitative of the invention.

ISOLATED means altered "by the hand of man" from its natural state; i.e., that, if it occurs in nature, it has been changed or removed from its original environment, or  
15 both.

For example, a naturally occurring polynucleotide or a polypeptide naturally present in a living animal in its natural state is not "isolated," but the same polynucleotide or polypeptide separated from the coexisting  
20 materials of its natural state is "isolated", as the term is employed herein. For example, with respect to polynucleotides, the term isolated means that it is separated from the chromosome and cell in which it naturally occurs.

25 As part of or following isolation, such polynucleotides can be joined to other polynucleotides, such as DNAs, for mutagenesis, to form fusion proteins, and for propagation or expression in a host, for instance. The isolated polynucleotides, alone or joined to other  
30 polynucleotides such as vectors, can be introduced into host cells, in culture or in whole organisms. When introduced into host cells in culture or in whole organisms, such DNAs still would be isolated, as the term is used herein, because they would not be in their

- 13 -

naturally occurring form or environment. Similarly, the polynucleotides and polypeptides may occur in a composition, such as media formulations, solutions for introduction of polynucleotides or polypeptides, for  
5 example, into cells, compositions or solutions for chemical or enzymatic reactions, for instance, which are not naturally occurring compositions, and, therein remain isolated polynucleotides or polypeptides within the meaning of that term as it is employed herein.

10           OLIGONUCLEOTIDE(S) refers to relatively short polynucleotides. Often the term refers to single-stranded deoxyribonucleotides, but it can refer as well to single-or double-stranded ribonucleotides, RNA:DNA hybrids and double-stranded DNAs, among others.

15           Oligonucleotides, such as single-stranded DNA probe oligonucleotides, often are synthesized by chemical methods, such as those implemented on automated oligonucleotide synthesizers. However, oligonucleotides can be made by a variety of other methods, including in  
20 vitro recombinant DNA-mediated techniques and by expression of DNAs in cells and organisms.

          Initially, chemically synthesized DNAs typically are obtained without a 5' phosphate. The 5' ends of such oligonucleotides are not substrates for phosphodiester bond  
25 formation by ligation reactions that employ DNA ligases typically used to form recombinant DNA molecules. Where ligation of such oligonucleotides is desired, a phosphate can be added by standard techniques, such as those that employ a kinase and ATP.

30           The 3' end of a chemically synthesized oligonucleotide generally has a free hydroxyl group and, in the presence of a ligase such as T4 DNA ligase, readily will form a phosphodiester bond with a 5' phosphate of another polynucleotide, such as another oligonucleotide.

35 As is well known, this reaction can be prevented

- 14 -

selectively, where desired, by removing the 5' phosphates of the other polynucleotide(s) prior to ligation.

POLYNUCLEOTIDE(S) generally refers to any polyribonucleotide or polydeoxribonucleotide and is inclusive of unmodified RNA or DNA as well as modified RNA or DNA. Thus, for instance, polynucleotides as used herein refers to, among other things, single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded or a mixture of single- and double-stranded regions. In addition, polynucleotide, as used herein, refers to triple-stranded regions comprising RNA or DNA or both RNA and DNA. The strands in such regions may be from the same molecule or from different molecules. The regions may include all of one or more of the molecules, but more typically involve only a region of some of the molecules. One of the molecules of a triple-helical region often is an oligonucleotide.

As used herein, the term polynucleotide is also inclusive of DNAs or RNAs as described above that contain one or more modified bases. Thus, DNAs or RNAs with backbones modified for stability or for other reasons are "polynucleotides" as that term is intended herein. Moreover, DNAs or RNAs comprising unusual bases, such as inosine, or modified bases, such as tritylated bases, to name just two examples, are polynucleotides as the term is used herein.

It will be appreciated that a great variety of modifications have been made to DNA and RNA that serve many useful purposes known to those of skill in the art. The term polynucleotide as it is employed herein embraces such chemically, enzymatically or metabolically modified forms

- 15 -

of polynucleotides, as well as chemical forms of DNA and RNA characteristic of viruses and cells, including simple and complex cells, inter alia.

POLYPEPTIDES, as used herein, includes all  
5 polypeptides as described below. The basic structure of polypeptides is well known and has been described in innumerable textbooks and other publications in the art. In this context, the term is used herein to refer to any peptide or protein comprising two or more amino acids  
10 joined to each other in a linear chain by peptide bonds. As used herein, the term refers to both short chains, which also commonly are referred to in the art as peptides, oligopeptides and oligomers, for example, and to longer chains, which generally are referred to in the art as  
15 proteins, of which there are many types. It will be appreciated that polypeptides often contain amino acids other than the 20 amino acids commonly referred to as the 20 naturally occurring amino acids, and that many amino acids, including the terminal amino acids, may be modified  
20 in a given polypeptide, either by natural processes such as processing and other post-translational modifications, or by chemical modification techniques which are well known to the art. Even the common modifications that occur naturally in polypeptides are too numerous to list  
25 exhaustively here, but they are well described in basic texts and in more detailed monographs, as well as in a voluminous research literature, and they are well known to those of skill in the art.

Modifications which may be present in polypeptides of  
30 the present invention include, to name an illustrative few, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or  
35 lipid derivative, covalent attachment of

- 16 -

phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cystine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor  
5 formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination.

10       Such modifications are well known to those of skill and have been described in great detail in the scientific literature. Several particularly common modifications including, but not limited to, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid  
15 residues, hydroxylation and ADP-ribosylation are described in most basic texts, such as, for instance PROTEINS STRUCTURE AND MOLECULAR PROPERTIES, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York (1993). Many detailed reviews are available on this subject, such as,  
20 for example, those provided by Wold, F., Posttranslational Protein Modifications: Perspectives and Prospects, pgs. 1-12 in POSTTRANSLATIONAL COVALENT MODIFICATION OF PROTEINS, B. C. Johnson, Ed., Academic Press, New York (1983); Seifter et al., Analysis for protein modifications and  
25 nonprotein cofactors, Meth. Enzymol. 182: 626-646 (1990) and Rattan et al., Protein Synthesis: Posttranslational Modifications and Aging, Ann. N.Y. Acad. Sci. 663: 48-62 (1992).

It will be appreciated that the polypeptides of the  
30 present invention are not always entirely linear. Instead, polypeptides may be branched as a result of ubiquitination, and they may be circular, with or without branching, generally as a result of posttranslation events including natural processing event and events brought about by human  
35 manipulation which do not occur naturally. Circular,

- 17 -

branched and branched circular polypeptides may be synthesized by non-translation natural processes and by entirely synthetic methods, as well.

Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. In fact, blockage of the amino and/or carboxyl group in a polypeptide by a covalent modification is common in naturally occurring and synthetic polypeptides and such modifications may be present in polypeptides of the present invention, as well. For instance, the amino terminal residue of polypeptides made in *E. coli*, prior to proteolytic processing, almost invariably will be N-formylmethionine.

The modifications that occur in a polypeptide often will be a function of how it is made. For polypeptides made by expressing a cloned gene in a host, for instance, the nature and extent of the modifications, in large part, will be determined by the host cell posttranslational modification capacity and the modification signals present in the polypeptide amino acid sequence. For instance, as is well known, glycosylation often does not occur in bacterial hosts such as *E. coli*. Accordingly, when glycosylation is desired, a polypeptide can be expressed in a glycosylating host, generally a eukaryotic cell. Insect cells often carry out the same posttranslational glycosylations as mammalian cells. Thus, insect cell expression systems have been developed to express efficiently mammalian proteins having native patterns of glycosylation, *inter alia*. Similar considerations apply to other modifications.

It will be appreciated that the same type of modification may be present in the same or varying degrees at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications.

- 18 -

In general, as used herein, the term polypeptide encompasses all such modifications, particularly those that are present in polypeptides synthesized by expressing a polynucleotide in a host cell.

5        VARIANT(S) of polynucleotides or polypeptides, as the term is used herein, are polynucleotides or polypeptides that differ from a reference polynucleotide or polypeptide, respectively.

10        With respect to variant polynucleotides, differences are generally limited so that the nucleotide sequences of the reference and the variant are closely similar overall and, in many regions, identical. Thus, changes in the nucleotide sequence of the variant may be silent. That is, they may not alter the amino acids encoded by the  
15        polynucleotide. Where alterations are limited to silent changes of this type a variant will encode a polypeptide with the same amino acid sequence as the reference. Alternatively, changes in the nucleotide sequence of the variant may alter the amino acid sequence of a polypeptide  
20        encoded by the reference polynucleotide. Such nucleotide changes may result in amino acid substitutions, additions, deletions, fusions and truncations in the polypeptide encoded by the reference sequence.

25        With respect to variant polypeptides, differences are generally limited so that the sequences of the reference and the variant are closely similar overall and, in many region, identical. For example, a variant and reference polypeptide may differ in amino acid sequence by one or more substitutions, additions, deletions, fusions and  
30        truncations, which may be present in any combination.

RECEPTOR MOLECULE, as used herein, refers to molecules which bind or interact specifically with LSG polypeptides of the present invention and is inclusive not only of classic receptors, which are preferred, but also  
35        other molecules that specifically bind to or interact with



- 19 -

polypeptides of the invention (which also may be referred to as "binding molecules" and "interaction molecules," respectively and as "LSG binding or interaction molecules". Binding between polypeptides of the invention and such

5 molecules, including receptor or binding or interaction molecules may be exclusive to polypeptides of the invention, which is very highly preferred, or it may be highly specific for polypeptides of the invention, which is highly preferred, or it may be highly specific to a group

10 of proteins that includes polypeptides of the invention, which is preferred, or it may be specific to several groups of proteins at least one of which includes polypeptides of the invention.

Receptors also may be non-naturally occurring, such as

15 as antibodies and antibody-derived reagents that bind to polypeptides of the invention.

#### DETAILED DESCRIPTION OF THE INVENTION

The present invention relates to novel lung specific polypeptides and polynucleotides, referred to herein as

20 LSGs, among other things, as described in greater detail below.

#### *Polynucleotides*

In accordance with one aspect of the present invention, there are provided isolated LSG polynucleotides

25 which encode LSG polypeptides.

Using the information provided herein, such as the polynucleotide sequences set out in SEQ ID NO:1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 37 or 38 a polynucleotide of the present invention

30 encoding a LSG may be obtained using standard cloning and screening procedures, such as those for cloning cDNAs using mRNA from cells of a human tumor as starting material.

Polynucleotides of the present invention may be in the form of RNA, such as mRNA, or in the form of DNA,

- 20 -

including, for instance, cDNA and genomic DNA obtained by cloning or produced by chemical synthetic techniques or by a combination thereof. The DNA may be double-stranded or single-stranded. Single-stranded DNA may be the coding strand, also known as the sense strand, or it may be the non-coding strand, also referred to as the anti-sense strand.

The coding sequence which encodes the polypeptides may be identical to the coding sequence of the polynucleotides of SEQ ID NO:1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 37 or 38. It also may be a polynucleotide with a different sequence, which, as a result of the redundancy (degeneracy) of the genetic code, encodes the same polypeptides as encoded by SEQ ID NO:1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 37 or 38.

Polynucleotides of the present invention, such as SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 37 or 38 which encode these polypeptides may comprise the coding sequence for the mature polypeptide by itself. Polynucleotides of the present invention may also comprise the coding sequence for the mature polypeptide and additional coding sequences such as those encoding a leader or secretory sequence such as a pre-, or pro- or prepro-protein sequence. Polynucleotides of the present invention may also comprise the coding sequence of the mature polypeptide, with or without the aforementioned additional coding sequences, together with additional, non-coding sequences. Examples of additional non-coding sequences which may be incorporated into the polynucleotide of the present invention include, but are not limited to, introns and non-coding 5' and 3' sequences such as transcribed, non-translated sequences that play a role in transcription, mRNA processing including, for example, splicing and polyadenylation signals, ribosome

- 21 -

binding and stability of mRNA, and additional coding sequence which codes for amino acids such as those which provide additional functionalities. Thus, for instance, the polypeptide may be fused to a marker sequence such as a peptide which facilitates purification of the fused polypeptide. In certain preferred embodiments of this aspect of the invention, the marker sequence is a hexahistidine peptide, such as the tag provided in the pQE vector (Qiagen, Inc.), among others, many of which are commercially available. As described in Gentz et al. (Proc. Natl. Acad. Sci., USA 86: 821-824 (1989)), for instance, hexahistidine provides for convenient purification of the fusion protein. The HA tag corresponds to an epitope derived of influenza hemagglutinin protein (Wilson et al., Cell 37: 767 (1984)).

In accordance with the foregoing, the term "polynucleotide encoding a polypeptide" as used herein encompasses polynucleotides which include a sequence encoding a polypeptide of the present invention, particularly SEQ ID NO:1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 37 or 38. Exemplary polypeptides encoded by the polynucleotides are depicted in SEQ ID NO: 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, or 56. The term encompasses polynucleotides that include a single continuous region or discontinuous regions encoding the polypeptide (for example, interrupted by introns) together with additional regions, that also may contain coding and/or non-coding sequences.

The present invention further relates to variants of the herein above described polynucleotides which encode for fragments, analogs and derivatives of the LSG polypeptides. A variant of the polynucleotide may be a naturally occurring variant such as a naturally occurring allelic variant, or it may be a variant that is not known to occur

- 22 -

naturally. Such non-naturally occurring variants of the polynucleotide may be made by mutagenesis techniques, including those applied to polynucleotides, cells or organisms.

5        Among variants in this regard are variants that differ from the aforementioned polynucleotides by nucleotide substitutions, deletions or additions. The substitutions, deletions or additions may involve one or more nucleotides. The variants may be altered in coding or  
10 non-coding regions or both. Alterations in the coding regions may produce conservative or non-conservative amino acid substitutions, deletions or additions.

      Among the particularly preferred embodiments of the invention in this regard are polynucleotides encoding  
15 polypeptides having the same amino acid sequence encoded by a LSG polynucleotide comprising SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 37 or 38; variants, analogs, derivatives and fragments thereof, and fragments of the variants, analogs and  
20 derivatives. Exemplary polypeptides encoded by these polynucleotides are depicted in SEQ ID NO: 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55 or 56. Further particularly preferred in this regard are LSG polynucleotides encoding polypeptide variants, analogs,  
25 derivatives and fragments, and variants, analogs and derivatives of the fragments, in which several, a few, 5 to 10, 1 to 5, 1 to 3, 2, 1 or no amino acid residues are substituted, deleted or added, in any combination. Especially preferred among these are silent substitutions,  
30 additions and deletions, which do not alter the properties and activities of the LSG. Also especially preferred in this regard are conservative substitutions. Most highly preferred are polynucleotides encoding polypeptides having the amino acid sequences as polypeptides encoded by SEQ ID

- 23 -

NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 37 or 38 without substitutions.

Further preferred embodiments of the invention are LSG polynucleotides that are at least 70% identical to a polynucleotide of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 37 or 38 and polynucleotides which are complementary to such polynucleotides. More preferred are LSG polynucleotides that comprise a region that is at least 80% identical to a polynucleotide of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 37 or 38. In this regard, LSG polynucleotides at least 90% identical to the same are particularly preferred, and among these particularly preferred LSG polynucleotides, those with at least 95% are especially preferred. Furthermore, those with at least 97% are highly preferred among those with at least 95%, and among these those with at least 98% and at least 99% are particularly highly preferred, with at least 99% being the most preferred.

Particularly preferred embodiments in this respect, moreover, are polynucleotides which encode polypeptides which retain substantially the same biological function or activity as the mature polypeptides encoded by a polynucleotide of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 37 or 38.

The present invention further relates to polynucleotides that hybridize to the herein above-described LSG sequences. In this regard, the present invention especially relates to polynucleotides which hybridize under stringent conditions to the herein above-described polynucleotides. As herein used, the term "stringent conditions" means hybridization will occur only if there is at least 95% and preferably at least 97% identity between the sequences.

- 24 -

As discussed additionally herein regarding polynucleotide assays of the invention, for instance, polynucleotides of the invention as described herein, may be used as a hybridization probe for cDNA and genomic DNA to isolate full-length cDNAs and genomic clones encoding LSGs and to isolate cDNA and genomic clones of other genes that have a high sequence similarity to these LSGs. Such probes generally will comprise at least 15 bases. Preferably, such probes will have at least 30 bases and may have at least 50 bases.

For example, the coding region of LSG of the present invention may be isolated by screening using an oligonucleotide probe synthesized from the known DNA sequence. A labeled oligonucleotide having a sequence complementary to that of a gene of the present invention is used to screen a library of human cDNA, genomic DNA or mRNA to determine which members of the library the probe hybridizes with.

The polynucleotides and polypeptides of the present invention may be employed as research reagents and materials for discovery of treatments and diagnostics to human disease, as further discussed herein relating to polynucleotide assays, *inter alia*.

The polynucleotides may encode a polypeptide which is the mature protein plus additional amino or carboxyl-terminal amino acids, or amino acids interior to the mature polypeptide (when the mature form has more than one polypeptide chain, for instance). Such sequences may play a role in processing of a protein from precursor to a mature form, may facilitate/protein trafficking, may prolong or shorten protein half-life or may facilitate manipulation of a protein for assay or production, among other things. As generally is the case *in situ*, the additional amino acids may be processed away from the mature protein by cellular enzymes.

- 25 -

A precursor protein having the mature form of the polypeptide fused to one or more prosequences may be an inactive form of the polypeptide. When prosequences are removed, such inactive precursors generally are activated.

5 Some or all of the prosequences may be removed before activation. Generally, such precursors are called proproteins.

In sum, a polynucleotide of the present invention may encode a mature protein, a mature protein plus a leader

10 sequence (which may be referred to as a preprotein), a precursor of a mature protein having one or more prosequences which are not the leader sequences of a preprotein, or a preproprotein, which is a precursor to a proprotein, having a leader sequence and one or more

15 prosequences, which generally are removed during processing steps that produce active and mature forms of the polypeptide.

### **Polypeptides**

The present invention further relates to LSG

20 polypeptides, preferably polypeptides encoded by a polynucleotide of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 37 or 38. Exemplary polypeptides are depicted in SEQ ID NO: 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55

25 or 56. The invention also relates to fragments, analogs and derivatives of these polypeptides. The terms "fragment," "derivative" and "analog" when referring to the polypeptides of the present invention means a polypeptide which retains essentially the same biological function or

30 activity as such polypeptides. Thus, an analog includes a proprotein which can be activated by cleavage of the proprotein portion to produce an active mature polypeptide.

The polypeptide of the present invention may be a recombinant polypeptide, a natural polypeptide or a .

- 26 -

synthetic polypeptide. In certain preferred embodiments it is a recombinant polypeptide.

The fragment, derivative or analog of a polypeptide of or the present invention may be (I) one in which one or  
5 more of the amino acid residues are substituted with a conserved or non-conserved amino acid residue (preferably a conserved amino acid residue) and such substituted amino acid residue may or may not be one encoded by the genetic code; (ii) one in which one or more of the amino acid  
10 residues includes a substituent group; (iii) one in which the mature polypeptide is fused with another compound, such as a compound to increase the half-life of the polypeptide (for example, polyethylene glycol); or (iv) one in which the additional amino acids are fused to the mature  
15 polypeptide, such as a leader or secretory sequence or a sequence which is employed for purification of the mature polypeptide or a proprotein sequence. Such fragments, derivatives and analogs are deemed to be within the scope of those skilled in the art from the teachings herein.

20 Among preferred variants are those that vary from a reference by conservative amino acid substitutions. Such substitutions are those that substitute a given amino acid in a polypeptide by another amino acid of like characteristics. Typically seen as conservative  
25 substitutions are the replacements, one for another, among the aliphatic amino acids Ala, Val, Leu and Ile; interchange of the hydroxyl residues Ser and Thr, exchange of the acidic residues Asp and Glu, substitution between the amide residues Asn and Gln, exchange of the basic  
30 residues Lys and Arg and replacements among the aromatic residues Phe, Tyr.

The polypeptides and polynucleotides of the present invention are preferably provided in an isolated form, and preferably are purified to homogeneity.



- 27 -

The polypeptides of the present invention include the polypeptides encoded by the polynucleotide of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 37 or 38 (in particular the mature polypeptide) as well as polypeptides which have at least 75% similarity (preferably at least 75% identity), more preferably at least 90% similarity (more preferably at least 90% identity), still more preferably at least 95% similarity (still more preferably at least 95% identity), to a polypeptide encoded by SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 37 or 38. Also included are portions of such polypeptides generally containing at least 30 amino acids and more preferably at least 50 amino acids. Exemplary polypeptides are depicted in SEQ ID NO: 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55 or 56.

As known in the art "similarity" between two polypeptides is determined by comparing the amino acid sequence and its conserved amino acid substitutes of one polypeptide sequence with that of a second polypeptide.

Fragments or portions of the polypeptides of the present invention may be employed for producing the corresponding full-length polypeptide by peptide synthesis; therefore, the fragments may be employed as intermediates for producing the full-length polypeptides. Fragments or portions of the polynucleotides of the present invention may be used to synthesize full-length polynucleotides of the present invention.

#### *Fragments*

Also among preferred embodiments of this aspect of the present invention are polypeptides comprising fragments of a polypeptide encoded by a polynucleotide of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 37 or 38. In this regard a fragment is a polypeptide having an amino acid sequence that entirely

- 28 -

is the same as part but not all of the amino acid sequence of the aforementioned LSG polypeptides and variants or derivatives thereof.

Such fragments may be "free-standing," i.e., not part  
5 of or fused to other amino acids or polypeptides, or they may be contained within a larger polypeptide of which they form a part or region. When contained within a larger polypeptide, the presently discussed fragments most preferably form a single continuous region. However,  
10 several fragments may be comprised within a single larger polypeptide. For instance, certain preferred embodiments relate to a fragment of a LSG polypeptide of the present comprised within a precursor polypeptide designed for expression in a host and having heterologous pre- and pro-  
15 polypeptide regions fused to the amino terminus of the LSG fragment and an additional region fused to the carboxyl terminus of the fragment. Therefore, fragments in one aspect of the meaning intended herein, refers to the portion or portions of a fusion polypeptide or fusion  
20 protein derived from a LSG polypeptide.

As representative examples of polypeptide fragments of the invention, there may be mentioned those which have from about 15 to about 139 amino acids. In this context "about" includes the particularly recited range and ranges  
25 larger or smaller by several, a few, 5, 4, 3, 2 or 1 amino acid at either extreme or at both extremes. Highly preferred in this regard are the recited ranges plus or minus as many as 5 amino acids at either or at both extremes. Particularly highly preferred are the recited  
30 ranges plus or minus as many as 3 amino acids at either or at both the recited extremes. Especially preferred are ranges plus or minus 1 amino acid at either or at both extremes or the recited ranges with no additions or deletions. Most highly preferred of all in this regard are  
35 fragments from about 15 to about 45 amino acids.

- 29 -

Among especially preferred fragments of the invention are truncation mutants of the LSG polypeptides. Truncation mutants include LSG polypeptides having an amino acid sequence encoded by a polynucleotide of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 37 or 38 or variants or derivatives thereof, except for deletion of a continuous series of residues (that is, a continuous region, part or portion) that includes the amino terminus, or a continuous series of residues that includes the carboxyl terminus or, as in double truncation mutants, deletion of two continuous series of residues, one including the amino terminus and one including the carboxyl terminus. Fragments having the size ranges set out herein also are preferred embodiments of truncation fragments, which are especially preferred among fragments generally.

Also preferred in this aspect of the invention are fragments characterized by structural or functional attributes of the LSG polypeptides of the present invention. Preferred embodiments of the invention in this regard include fragments that comprise alpha-helix and alpha-helix forming regions ("alpha-regions"), beta-sheet and beta-sheet-forming regions ("beta-regions"), turn and turn-forming regions ("turn-regions"), coil and coil-forming regions ("coil-regions"), hydrophilic regions, hydrophobic regions, alpha amphipathic regions, beta amphipathic regions, flexible regions, surface-forming regions and high antigenic index regions of the LSG polypeptides of the present invention. Regions of the aforementioned types are identified routinely by analysis of the amino acid sequences encoded by the polynucleotides of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 37 or 38. Preferred regions include Garnier-Robson alpha-regions, beta-regions, turn-regions and coil-regions, Chou-Fasman alpha-regions,

- 30 -

beta-regions and turn-regions, Kyte-Doolittle hydrophilic regions and hydrophilic regions, Eisenberg alpha and beta amphipathic regions, Karplus-Schulz flexible regions, Emini surface-forming regions and Jameson-Wolf high antigenic index regions. Among highly preferred fragments in this regard are those that comprise regions of LSGs that combine several structural features, such as several of the features set out above. In this regard, the regions defined by selected residues of a LSG polypeptide which all are characterized by amino acid compositions highly characteristic of turn-regions, hydrophilic regions, flexible-regions, surface-forming regions, and high antigenic index-regions, are especially highly preferred regions. Such regions may be comprised within a larger polypeptide or may be by themselves a preferred fragment of the present invention, as discussed above. It will be appreciated that the term "about" as used in this paragraph has the meaning set out above regarding fragments in general.

Further preferred regions are those that mediate activities of LSG polypeptides. Most highly preferred in this regard are fragments that have a chemical, biological or other activity of a LSG polypeptide, including those with a similar activity or an improved activity, or with a decreased undesirable activity. Highly preferred in this regard are fragments that contain regions that are homologs in sequence, or in position, or in both sequence and to active regions of related polypeptides, and which include lung specific-binding proteins. Among particularly preferred fragments in these regards are truncation mutants, as discussed above.

It will be appreciated that the invention also relates to polynucleotides encoding the aforementioned fragments, polynucleotides that hybridize to polynucleotides encoding the fragments, particularly those

- 31 -

that hybridize under stringent conditions, and polynucleotides such as PCR primers for amplifying polynucleotides that encode the fragments. In these regards, preferred polynucleotides are those that correspond to the preferred fragments, as discussed above.

#### *Fusion Proteins*

In one embodiment of the present invention, the LSG polypeptides of the present invention are preferably fused to other proteins. These fusion proteins can be used for a variety of applications. For example, fusion of the present polypeptides to His-tag, HA-tag, protein A, IgG domains, and maltose binding protein facilitates purification. (See also EP A 394,827; Traunecker, et al., Nature 331: 84-86 (1988)) Similarly, fusion to IgG-1, IgG-3, and albumin increases the halflife time in vivo. Nuclear localization signals fused to the polypeptides of the present invention can target the protein to a specific subcellular localization, while covalent heterodimer or homodimers can increase or decrease the activity of a fusion protein. Fusion proteins can also create chimeric molecules having more than one function. Finally, fusion proteins can increase solubility and/or stability of the fused protein compared to the non-fused protein. All of these types of fusion proteins described above can be made in accordance with well known protocols.

For example, a LSG polypeptide can be fused to an IgG molecule via the following protocol. Briefly, the human Fc portion of the IgG molecule is PCR amplified using primers that span the 5' and 3' ends of the sequence. These primers also have convenient restriction enzyme sites that facilitate cloning into an expression vector, preferably a mammalian expression vector. For example, if pC4 (Accession No. 209646) is used, the human Fc portion can be ligated into the BamHI cloning site. In this protocol, the 3' BamHI site must be destroyed. Next, the vector

- 32 -

containing the human Fc portion is re-restricted with BamHI thereby linearizing the vector, and a LSG polynucleotide of the present invention is ligated into this BamHI site. It is preferred that the polynucleotide is cloned without a stop codon, otherwise a fusion protein will not be produced.

If the naturally occurring signal sequence is used to produce the secreted protein, pC4 does not need a second signal peptide. Alternatively, if the naturally occurring signal sequence is not used, the vector can be modified to include a heterologous signal sequence. (See, e. g., WO 96/34891.)

#### *Diagnostic Assays*

The present invention also relates to diagnostic assays and methods, both quantitative and qualitative for detecting, diagnosing, monitoring, staging and prognosticating cancers by comparing levels of LSG in a human patient with those of LSG in a normal human control. For purposes of the present invention, what is meant by LSG levels is, among other things, native protein expressed by a gene comprising the polynucleotide sequence of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 37 or 38. Exemplary polypeptides encoded by these polynucleotides are depicted in SEQ ID NO: 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55 or 56. By "LSG" it is also meant herein polynucleotides which, due to degeneracy in genetic coding, comprise variations in nucleotide sequence as compared to SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 37 or 38 but which still encode the same protein. The native protein being detected may be whole, a breakdown product, a complex of molecules or chemically modified. In the alternative, what is meant by LSG as used herein, means the native mRNA encoded by a polynucleotide sequence of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7,

- 33 -

8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, or  
22, or a contig of SEQ ID NO:19 or 21, depicted as SEQ ID  
NO: 37 or 38, respectively, levels of the gene comprising  
the polynucleotide sequence of SEQ ID NO: 1, 2, 3, 4, 5, 6,  
5 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, or  
22, or levels of a polynucleotide which is capable of  
hybridizing under stringent conditions to the antisense  
sequence of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11,  
12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 37 or 38. Such  
10 levels are preferably determined in at least one of cells,  
tissues and/or bodily fluids, including determination of  
normal and abnormal levels. Thus, for instance, a  
diagnostic assay in accordance with the invention for  
diagnosing overexpression of LSG protein compared to normal  
15 control bodily fluids, cells, or tissue samples may be used  
to diagnose the presence of lung cancer.

All the methods of the present invention may  
optionally include determining the levels of other cancer  
markers as well as LSG. Other cancer markers, in addition  
20 to LSG, useful in the present invention will depend on the  
cancer being tested and are known to those of skill in the  
art.

The present invention provides methods for diagnosing  
the presence of lung cancer by analyzing for changes in  
25 levels of LSG in cells, tissues or bodily fluids compared  
with levels of LSG in cells, tissues or bodily fluids of  
preferably the same type from a normal human control,  
wherein an increase in levels of LSG in the patient versus  
the normal human control is associated with the presence of  
30 lung cancer.

Without limiting the instant invention, typically,  
for a quantitative diagnostic assay a positive result  
indicating the patient being tested has cancer is one in  
which cells, tissues or bodily fluid levels of the cancer  
35 marker, such as LSG, are at least two times higher, and

- 34 -

most preferably are at least five times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control.

The present invention also provides a method of  
5 diagnosing metastatic lung cancer in a patient having lung cancer which has not yet metastasized for the onset of metastasis. In the method of the present invention, a human cancer patient suspected of having lung cancer which may have metastasized (but which was not previously known  
10 to have metastasized) is identified. This is accomplished by a variety of means known to those of skill in the art.

In the present invention, determining the presence of LSG levels in cells, tissues or bodily fluid, is particularly useful for discriminating between lung cancer  
15 which has not metastasized and lung cancer which has metastasized. Existing techniques have difficulty discriminating between lung cancer which has metastasized and lung cancer which has not metastasized and proper treatment selection is often dependent upon such knowledge.

20 In the present invention, the cancer marker levels measured in such cells, tissues or bodily fluid is LSG, and are compared with levels of LSG in preferably the same cells, tissue or bodily fluid type of a normal human control. That is, if the cancer marker being observed is  
25 just LSG in serum, this level is preferably compared with the level of LSG in serum of a normal human control. An increase in the LSG in the patient versus the normal human control is associated with lung cancer which has metastasized.

30 Without limiting the instant invention, typically, for a quantitative diagnostic assay a positive result indicating the cancer in the patient being tested or monitored has metastasized is one in which cells, tissues or bodily fluid levels of the cancer marker, such as LSG,  
35 are at least two times higher, and most preferably are at



- 35 -

least five times higher, than in preferably the same cells, tissues or bodily fluid of a normal patient.

Normal human control as used herein includes a human patient without cancer and/or non cancerous samples from  
5 the patient; in the methods for diagnosing or monitoring for metastasis, normal human control may preferably also include samples from a human patient that is determined by reliable methods to have lung cancer which has not metastasized.

#### 10 *Staging*

The invention also provides a method of staging lung cancer in a human patient. The method comprises identifying a human patient having such cancer and analyzing cells, tissues or bodily fluid from such human  
15 patient for LSG. The LSG levels determined in the patient are then compared with levels of LSG in preferably the same cells, tissues or bodily fluid type of a normal human control, wherein an increase in LSG levels in the human patient versus the normal human control is associated with  
20 a cancer which is progressing and a decrease in the levels of LSG (but still increased over true normal levels) is associated with a cancer which is regressing or in remission.

#### *Monitoring*

Further provided is a method of monitoring lung  
25 cancer in a human patient having such cancer for the onset of metastasis. The method comprises identifying a human patient having such cancer that is not known to have metastasized; periodically analyzing cells, tissues or  
30 bodily fluid from such human patient for LSG; and comparing the LSG levels determined in the human patient with levels of LSG in preferably the same cells, tissues or bodily fluid type of a normal human control, wherein an increase in LSG levels in the human patient versus the normal human  
35 control is associated with a cancer which has metastasized.

- 36 -

In this method, normal human control samples may also include prior patient samples.

Further provided by this invention is a method of monitoring the change in stage of lung cancer in a human patient having such cancer. The method comprises identifying a human patient having such cancer; periodically analyzing cells, tissues or bodily fluid from such human patient for LSG; and comparing the LSG levels determined in the human patient with levels of LSG in preferably the same cells, tissues or bodily fluid type of a normal human control, wherein an increase in LSG levels in the human patient versus the normal human control is associated with a cancer which is progressing in stage and a decrease in the levels of LSG is associated with a cancer which is regressing in stage or in remission. In this method, normal human control samples may also include prior patient samples.

Monitoring a patient for onset of metastasis is periodic and preferably done on a quarterly basis. However, this may be done more or less frequently depending on the cancer, the particular patient, and the stage of the cancer.

#### *Prognostic Testing and Clinical Trial Monitoring*

The methods described herein can further be utilized as prognostic assays to identify subjects having or at risk of developing a disease or disorder associated with increased levels of LSG. The present invention provides a method in which a test sample is obtained from a human patient and LSG is detected. The presence of higher LSG levels as compared to normal human controls is diagnostic for the human patient being at risk for developing cancer, particularly lung cancer.

The effectiveness of therapeutic agents to decrease expression or activity of the LSGs of the invention can also be monitored by analyzing levels of expression of the

- 37 -

LSGs in a human patient in clinical trials or in *in vitro* screening assays such as in human cells. In this way, the gene expression pattern can serve as a marker, indicative of the physiological response of the human patient, or  
5 cells as the case may be, to the agent being tested.

*Detection of genetic lesions or mutations*

The methods of the present invention can also be used to detect genetic lesions or mutations in LSG, thereby determining if a human with the genetic lesion is at risk  
10 for lung cancer or has lung cancer. Genetic lesions can be detected, for example, by ascertaining the existence of a deletion and/or addition and/or substitution of one or more nucleotides from the LSGs of this invention, a chromosomal rearrangement of LSG, aberrant modification of LSG (such as  
15 of the methylation pattern of the genomic DNA), the presence of a non-wild type splicing pattern of a mRNA transcript of LSG, allelic loss of LSG, and/or inappropriate post-translational modification of LSG protein. Methods to detect such lesions in the LSG of this  
20 invention are known to those of skill in the art.

For example, in one embodiment, alterations in a gene corresponding to a LSG polynucleotide of the present invention are determined via isolation of RNA from entire families or individual patients presenting with a phenotype  
25 of interest (such as a disease) is be isolated. cDNA is then generated from these RNA samples using protocols known in the art. See, e.g. Sambrook et al. (MOLECULAR CLONING: A LABORATORY MANUAL, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989)), which is  
30 illustrative of the many laboratory manuals that detail these techniques. The cDNA is then used as a template for PCR, employing primers surrounding regions of interest in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 37 or 38. PCR conditions  
35 typically consist of 35 cycles at 95°C for 30 seconds; 60-

- 38 -

120 seconds at 52-58°C; and 60-120 seconds at 70°C, using buffer solutions described in Sidransky, D., et al., Science 252: 706 (1991). PCR products are sequenced using primers labeled at their 5' end with T4 polynucleotide kinase, employing SequiTherm Polymerase (Epicentre Technologies). The intron-exon borders of selected exons are also determined and genomic PCR products analyzed to confirm the results. PCR products harboring suspected mutations are then cloned and sequenced to validate the results of the direct sequencing. PCR products are cloned into T-tailed vectors as described in Holton, T. A. and Graham, M. W., Nucleic Acids Research, 19 : 1156 (1991) and sequenced with T7 polymerase (United States Biochemical). Affected individuals are identified by mutations not present in unaffected individuals.

Genomic rearrangements can also be observed as a method of determining alterations in a gene corresponding to a polynucleotide. In this method, genomic clones are nick-translated with digoxigenin deoxy-uridine 5'-triphosphate (Boehringer Mannheim), and FISH is performed as described in Johnson, C. et al., Methods Cell Biol. 35: 73-99 (1991). Hybridization with a labeled probe is carried out using a vast excess of human DNA for specific hybridization to the corresponding genomic locus. Chromosomes are counterstained with 4,6-diamino-2-phenylidole and propidium iodide, producing a combination of C-and R-bands. Aligned images for precise mapping are obtained using a triple-band filter set (Chroma Technology, Brattleboro, VT) in combination with a cooled charge-coupled device camera (Photometrics, Tucson, AZ) and variable excitation wavelength filters (Johnson et al., Genet. Anal. Tech. Appl., 8: 75 (1991)). Image collection, analysis and chromosomal fractional length measurements are performed using the ISee Graphical Program System (Inovision Corporation, Durham, NC). Chromosome

- 39 -

alterations of the genomic region hybridized by the probe are identified as insertions, deletions, and translocations. These alterations are used as a diagnostic marker for an associated disease.

## 5 Assay Techniques

Assay techniques that can be used to determine levels of gene expression (including protein levels), such as LSG of the present invention, in a sample derived from a patient are well known to those of skill in the art. Such  
10 assay methods include, without limitation, radioimmunoassays, reverse transcriptase PCR (RT-PCR) assays, immunohistochemistry assays, *in situ* hybridization assays, competitive-binding assays, Western Blot analyses, ELISA assays and proteomic approaches: two-dimensional gel  
15 electrophoresis (2D electrophoresis) and non-gel based approaches such as mass spectrometry or protein interaction profiling. Among these, ELISAs are frequently preferred to diagnose a gene's expressed protein in biological fluids.

An ELISA assay initially comprises preparing an  
20 antibody, if not readily available from a commercial source, specific to LSG, preferably a monoclonal antibody. In addition a reporter antibody generally is prepared which binds specifically to LSG. The reporter antibody is attached to a detectable reagent such as radioactive,  
25 fluorescent or enzymatic reagent, for example horseradish peroxidase enzyme or alkaline phosphatase.

To carry out the ELISA, antibody specific to LSG is incubated on a solid support, e.g. a polystyrene dish, that binds the antibody. Any free protein binding sites on the  
30 dish are then covered by incubating with a non-specific protein such as bovine serum albumin. Next, the sample to be analyzed is incubated in the dish, during which time LSG binds to the specific antibody attached to the polystyrene dish. Unbound sample is washed out with buffer. A reporter  
35 antibody specifically directed to LSG and linked to a

- 40 -

detectable reagent such as horseradish peroxidase is placed in the dish resulting in binding of the reporter antibody to any monoclonal antibody bound to LSG. Unattached reporter antibody is then washed out. Reagents for  
5 peroxidase activity, including a colorimetric substrate are then added to the dish. Immobilized peroxidase, linked to LSG antibodies, produces a colored reaction product. The amount of color developed in a given time period is proportional to the amount of LSG protein present in the  
10 sample. Quantitative results typically are obtained by reference to a standard curve.

A competition assay can also be employed wherein antibodies specific to LSG are attached to a solid support and labeled LSG and a sample derived from the host are  
15 passed over the solid support. The amount of label detected which is attached to the solid support can be correlated to a quantity of LSG in the sample.

Using all or a portion of a nucleic acid sequence of LSG of the present invention as a hybridization probe,  
20 nucleic acid methods can also be used to detect LSG mRNA as a marker for lung cancer. Polymerase chain reaction (PCR) and other nucleic acid methods, such as ligase chain reaction (LCR) and nucleic acid sequence based amplification (NASBA), can be used to detect malignant  
25 cells for diagnosis and monitoring of various malignancies. For example, reverse-transcriptase PCR (RT-PCR) is a powerful technique which can be used to detect the presence of a specific mRNA population in a complex mixture of thousands of other mRNA species. In RT-PCR, an mRNA  
30 species is first reverse transcribed to complementary DNA (cDNA) with use of the enzyme reverse transcriptase; the cDNA is then amplified as in a standard PCR reaction. RT-PCR can thus reveal by amplification the presence of a single species of mRNA. Accordingly, if the mRNA is highly

- 41 -

specific for the cell that produces it, RT-PCR can be used to identify the presence of a specific type of cell.

Hybridization to clones or oligonucleotides arrayed on a solid support (i.e. gridding) can be used to both  
5 detect the expression of and quantitate the level of expression of that gene. In this approach, a cDNA encoding the LSG gene is fixed to a substrate. The substrate may be of any suitable type including but not limited to glass, nitrocellulose, nylon or plastic. At least a portion of  
10 the DNA encoding the LSG gene is attached to the substrate and then incubated with the analyte, which may be RNA or a complementary DNA (cDNA) copy of the RNA, isolated from the tissue of interest. Hybridization between the substrate bound DNA and the analyte can be detected and quantitated  
15 by several means including but not limited to radioactive labeling or fluorescence labeling of the analyte or a secondary molecule designed to detect the hybrid. Quantitation of the level of gene expression can be done by comparison of the intensity of the signal from the analyte  
20 compared with that determined from known standards. The standards can be obtained by *in vitro* transcription of the target gene, quantitating the yield, and then using that material to generate a standard curve.

Of the proteomic approaches, 2D electrophoresis is a  
25 technique well known to those in the art. Isolation of individual proteins from a sample such as serum is accomplished using sequential separation of proteins by different characteristics usually on polyacrylamide gels. First, proteins are separated by size using an electric  
30 current. The current acts uniformly on all proteins, so smaller proteins move farther on the gel than larger proteins. The second dimension applies a current perpendicular to the first and separates proteins not on the basis of size but on the specific electric charge  
35 carried by each protein. Since no two proteins with

- 42 -

different sequences are identical on the basis of both size and charge, the result of a 2D separation is a square gel in which each protein occupies a unique spot. Analysis of the spots with chemical or antibody probes, or subsequent  
5 protein microsequencing can reveal the relative abundance of a given protein and the identity of the proteins in the sample.

The above tests can be carried out on samples derived from a variety of cells, bodily fluids and/or tissue  
10 extracts such as homogenates or solubilized tissue obtained from a patient. Tissue extracts are obtained routinely from tissue biopsy and autopsy material. Bodily fluids useful in the present invention include blood, urine, saliva or any other bodily secretion or derivative thereof.  
15 By blood it is meant to include whole blood, plasma, serum or any derivative of blood.

#### *In Vivo Targeting of LSG/Lung Cancer Therapy*

Identification of this LSG is also useful in the rational design of new therapeutics for imaging and  
20 treating cancers, and in particular lung cancer. For example, in one embodiment, antibodies which specifically bind to LSG can be raised and used *in vivo* in patients suspected of suffering from lung cancer. Antibodies which specifically bind LSG can be injected into a patient  
25 suspected of having lung cancer for diagnostic and/or therapeutic purposes. Thus, another aspect of the present invention provides for a method for preventing the onset and treatment of lung cancer in a human patient in need of such treatment by administering to the patient an effective  
30 amount of antibody. By "effective amount" it is meant the amount or concentration of antibody needed to bind to the target antigens expressed on the tumor to cause tumor shrinkage for surgical removal, or disappearance of the tumor. The binding of the antibody to the overexpressed  
35 LSG is believed to cause the death of the cancer cell



- 43 -

expressing such LSG. The preparation and use of antibodies for *in vivo* diagnosis and treatment is well known in the art. For example, antibody-chelators labeled with Indium-111 have been described for use in the

5 radioimmunoscintigraphic imaging of carcinoembryonic antigen expressing tumors (Sumerdon et al. Nucl. Med. Biol. 1990 17:247-254). In particular, these antibody-chelators have been used in detecting tumors in patients suspected of having recurrent colorectal cancer (Griffin et al. J. Clin. 10 Onc. 1991 9:631-640). Antibodies with paramagnetic ions as labels for use in magnetic resonance imaging have also been described (Lauffer, R.B. Magnetic Resonance in Medicine 1991 22:339-342). Antibodies directed against LSG can be used in a similar manner. Labeled antibodies which

15 specifically bind LSG can be injected into patients suspected of having lung cancer for the purpose of diagnosing or staging of the disease status of the patient. The label used will be selected in accordance with the imaging modality to be used. For example, radioactive

20 labels such as Indium-111, Technetium-99m or Iodine-131 can be used for planar scans or single photon emission computed tomography (SPECT). Positron emitting labels such as Fluorine-19 can be used in positron emission tomography. Paramagnetic ions such as Gadolinium (III) or Manganese (II)

25 can be used in magnetic resonance imaging (MRI). Presence of the label, as compared to imaging of normal tissue, permits determination of the spread of the cancer. The amount of label within an organ or tissue also allows determination of the presence or absence of cancer in that

30 organ or tissue.

Antibodies which can be used in *in vivo* methods include polyclonal, monoclonal and omniclonal antibodies and antibodies prepared via molecular biology techniques. Antibody fragments and aptamers and single-stranded

35 oligonucleotides such as those derived from an *in vitro*

- 44 -

evolution protocol referred to as SELEX and well known to those skilled in the art can also be used.

#### *Screening Assays*

The present invention also provides methods for  
5 identifying modulators which bind to LSG protein or have a  
modulatory effect on the expression or activity of LSG  
protein. Modulators which decrease the expression or  
activity of LSG protein are believed to be useful in  
treating lung cancer. Such screening assays are known to  
10 those of skill in the art and include, without limitation,  
cell-based assays and cell free assays.

Small molecules predicted via computer imaging to  
specifically bind to regions of LSG can also be designed,  
synthesized and tested for use in the imaging and treatment  
15 of lung cancer. Further, libraries of molecules can be  
screened for potential anticancer agents by assessing the  
ability of the molecule to bind to the LSGs identified  
herein. Molecules identified in the library as being  
capable of binding to LSG are key candidates for further  
20 evaluation for use in the treatment of lung cancer. In a  
preferred embodiment, these molecules will downregulate  
expression and/or activity of LSG in cells.

#### *Adoptive Immunotherapy and Vaccines*

Adoptive immunotherapy of cancer refers to a  
25 therapeutic approach in which immune cells with an  
antitumor reactivity are administered to a tumor-bearing  
host, with the aim that the cells mediate either directly  
or indirectly, the regression of an established tumor.  
Transfusion of lymphocytes, particularly T lymphocytes,  
30 falls into this category and investigators at the National  
Cancer Institute (NCI) have used autologous reinfusion of  
peripheral blood lymphocytes or tumor-infiltrating  
lymphocytes (TIL), T cell cultures from biopsies of  
subcutaneous lymph nodules, to treat several human cancers  
35 (Rosenberg, S. A., U.S. Patent No. 4,690,914, issued Sep.

- 45 -

1, 1987; Rosenberg, S. A., et al., 1988, N. England J. Med. 319:1676-1680).

The present invention relates to compositions and methods of adoptive immunotherapy for the prevention and/or treatment of primary and metastatic lung cancer in humans using macrophages sensitized to the antigenic LSG molecules, with or without non-covalent complexes of heat shock protein (hsp). Antigenicity or immunogenicity of the LSG is readily confirmed by the ability of the LSG protein or a fragment thereof to raise antibodies or educate naive effector cells, which in turn lyse target cells expressing the antigen (or epitope).

Cancer cells are, by definition, abnormal and contain proteins which should be recognized by the immune system as foreign since they are not present in normal tissues. However, the immune system often seems to ignore this abnormality and fails to attack tumors. The foreign LSG proteins that are produced by the cancer cells can be used to reveal their presence. The LSG is broken into short fragments, called tumor antigens, which are displayed on the surface of the cell. These tumor antigens are held or presented on the cell surface by molecules called MHC, of which there are two types: class I and II. Tumor antigens in association with MHC class I molecules are recognized by cytotoxic T cells while antigen-MHC class II complexes are recognized by a second subset of T cells called helper cells. These cells secrete cytokines which slow or stop tumor growth and help another type of white blood cell, B cells, to make antibodies against the tumor cells.

In adoptive immunotherapy, T cells or other antigen presenting cells (APCs) are stimulated outside the body (ex vivo), using the tumor specific LSG antigen. The stimulated cells are then reinfused into the patient where they attack the cancerous cells. Research has shown that using both cytotoxic and helper T cells is far more

- 46 -

effective than using either subset alone. Additionally, the LSG antigen may be complexed with heat shock proteins to stimulate the APCs as described in U.S. Patent No. 5,985,270.

5       The APCs can be selected from among those antigen presenting cells known in the art including, but not limited to, macrophages, dendritic cells, B lymphocytes, and a combination thereof, and are preferably macrophages. In a preferred use, wherein cells are autologous to the individual, autologous immune cells such as lymphocytes, 10 macrophages or other APCs are used to circumvent the issue of whom to select as the donor of the immune cells for adoptive transfer. Another problem circumvented by use of autologous immune cells is graft versus host disease which 15 can be fatal if unsuccessfully treated.

In adoptive immunotherapy with gene therapy, DNA of the LSG can be introduced into effector cells similarly as in conventional gene therapy. This can enhance the cytotoxicity of the effector cells to tumor cells as they 20 have been manipulated to produce the antigenic protein resulting in improvement of the adoptive immunotherapy.

LSG antigens of this invention are also useful as components of lung cancer vaccines. The vaccine comprises an immunogenically stimulatory amount of a LSG antigen. 25 Immunogenically stimulatory amount refers to that amount of antigen that is able to invoke the desired immune response in the recipient for the amelioration, or treatment of lung cancer. Effective amounts may be determined empirically by standard procedures well known to those skilled in the art.

30       The LSG antigen may be provided in any one of a number of vaccine formulations which are designed to induce the desired type of immune response, e.g., antibody and/or cell mediated. Such formulations are known in the art and include, but are not limited to, formulations such as those 35 described in U.S. Patent 5,585,103. Vaccine formulations

- 47 -

of the present invention used to stimulate immune responses can also include pharmaceutically acceptable adjuvants.

**Vectors, host cells, expression**

The present invention also relates to vectors which  
5 include polynucleotides of the present invention, host cells which are genetically engineered with vectors of the invention and the production of polypeptides of the invention by recombinant techniques.

Host cells can be genetically engineered to  
10 incorporate LSG polynucleotides and express LSG polypeptides of the present invention. For instance, LSG polynucleotides may be introduced into host cells using well known techniques of infection, transduction, transfection, transvection and transformation. The LSG  
15 polynucleotides may be introduced alone or with other polynucleotides. Such other polynucleotides may be introduced independently, co-introduced or introduced joined to the LSG polynucleotides of the invention.

For example, LSG polynucleotides of the invention may  
20 be transfected into host cells with another, separate, polynucleotide encoding a selectable marker, using standard techniques for co-transfection and selection in, for instance, mammalian cells. In this case, the polynucleotides generally will be stably incorporated into  
25 the host cell genome.

Alternatively, the LSG polynucleotide may be joined to a vector containing a selectable marker for propagation in a host. The vector construct may be introduced into host cells by the aforementioned techniques. Generally, a  
30 plasmid vector is introduced as DNA in a precipitate, such as a calcium phosphate precipitate, or in a complex with a charged lipid. Electroporation also may be used to introduce LSG polynucleotides into a host. If the vector is a virus, it may be packaged in vitro or introduced into  
35 a packaging cell and the packaged virus may be transduced

- 48 -

into cells. A wide variety of well known techniques conducted routinely by those of skill in the art are suitable for making LSG polynucleotides and for introducing LSG polynucleotides into cells in accordance with this  
5 aspect of the invention. Such techniques are reviewed at length in reference texts such as Sambrook et al., previously cited herein.

Vectors which may be used in the present invention include, for example, plasmid vectors, single- or double-  
10 stranded phage vectors, and single- or double-stranded RNA or DNA viral vectors. Such vectors may be introduced into cells as polynucleotides, preferably DNA, by well known techniques for introducing DNA and RNA into cells. The vectors, in the case of phage and viral vectors, also may  
15 be and preferably are introduced into cells as packaged or encapsidated virus by well known techniques for infection and transduction. Viral vectors may be replication competent or replication defective. In the latter case viral propagation generally will occur only in  
20 complementing host cells.

Preferred vectors for expression of polynucleotides and polypeptides of the present invention include, but are not limited to, vectors comprising cis-acting control regions effective for expression in a host operatively  
25 linked to the polynucleotide to be expressed. Appropriate trans-acting factors either are supplied by the host, supplied by a complementing vector or supplied by the vector itself upon introduction into the host.

In certain preferred embodiments in this regard, the  
30 vectors provide for specific expression. Such specific expression may be inducible expression or expression only in certain types of cells or both inducible and cell-specific. Particularly preferred among inducible vectors are vectors that can be induced to express by environmental  
35 factors that are easy to manipulate, such as temperature

- 49 -

and nutrient additives. A variety of vectors suitable to this aspect of the invention, including constitutive and inducible expression vectors for use in prokaryotic and eukaryotic hosts, are well known and employed routinely by those of skill in the art.

The engineered host cells can be cultured in conventional nutrient media which may be modified as appropriate for, *inter alia*, activating promoters, selecting transformants or amplifying genes. Culture conditions such as temperature, pH and the like, previously used with the host cell selected for expression, generally will be suitable for expression of LSG polypeptides of the present invention.

A great variety of expression vectors can be used to express LSG polypeptides of the invention. Such vectors include chromosomal, episomal and virus-derived vectors. Vectors may be derived from bacterial plasmids, from bacteriophage, from yeast episomes, from yeast chromosomal elements, from viruses such as baculoviruses, papova viruses, such as SV40, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses and retroviruses, and from combinations thereof such as those derived from plasmid and bacteriophage genetic elements, such as cosmids and phagemids. All may be used for expression in accordance with this aspect of the present invention. Generally, any vector suitable to maintain, propagate or express polynucleotides to express a polypeptide in a host may be used for expression in this regard.

The appropriate DNA sequence may be inserted into the vector by any of a variety of well-known and routine techniques. In general, a DNA sequence for expression is joined to an expression vector by cleaving the DNA sequence and the expression vector with one or more restriction endonucleases and then joining the restriction fragments together using T4 DNA ligase. Procedures for restriction

- 50 -

and ligation that can be used to this end are well known and routine to those of skill. Suitable procedures in this regard, and for constructing expression vectors using alternative techniques, which also are well known and  
5 routine to those skill, are set forth in great detail in Sambrook et al. cited elsewhere herein.

The DNA sequence in the expression vector is operatively linked to appropriate expression control sequence(s), including, for instance, a promoter to direct  
10 mRNA transcription. Representative promoters include the phage lambda PL promoter, the *E. coli* lac, trp and tac promoters, the SV40 early and late promoters, and promoters of retroviral LTRs, to name just a few of the well-known promoters. It will be understood that numerous promoters  
15 not mentioned are also suitable for use in this aspect of the invention and are well known and readily may be employed by those of skill in the manner illustrated by the discussion and the examples herein.

In general, expression constructs will contain sites  
20 for transcription initiation and termination, and, in the transcribed region, a ribosome binding site for translation. The coding portion of the mature transcripts expressed by the constructs will include a translation initiating AUG at the beginning and a termination codon  
25 appropriately positioned at the end of the polypeptide to be translated.

In addition, the constructs may contain control regions that regulate as well as engender expression. Generally, in accordance with many commonly practiced  
30 procedures, such regions will operate by controlling transcription, such as repressor binding sites and enhancers, among others.

Vectors for propagation and expression generally will include selectable markers. Such markers also may be  
35 suitable for amplification or the vectors may contain



- 51 -

additional markers for this purpose. In this regard, the expression vectors preferably contain one or more selectable marker genes to provide a phenotypic trait for selection of transformed host cells. Preferred markers  
5 include dihydrofolate reductase or neomycin resistance for eukaryotic cell culture, and tetracycline or ampicillin resistance genes for culturing in *E. coli* and other bacteria.

The vector containing the appropriate DNA sequence as  
10 described elsewhere herein, as well as an appropriate promoter, and other appropriate control sequences, may be introduced into an appropriate host using a variety of well known techniques suitable to expression therein of a desired polypeptide. Representative examples of  
15 appropriate hosts include bacterial cells, such as *E. coli*, *Streptomyces* and *Salmonella typhimurium* cells; fungal cells, such as yeast cells; insect cells such as *Drosophila* S2 and *Spodoptera Sf9* cells; animal cells such as CHO, COS and Bowes melanoma cells; and plant cells. Hosts for a  
20 great variety of expression constructs are well known, and those of skill will be enabled by the present disclosure readily to select a host for expressing a LSG polypeptide in accordance with this aspect of the present invention.

More particularly, the present invention also  
25 includes recombinant constructs, such as expression constructs, comprising one or more of the sequences described above. The constructs comprise a vector, such as a plasmid or viral vector, into which such LSG sequence of the invention has been inserted. The sequence may be  
30 inserted in a forward or reverse orientation. In certain preferred embodiments in this regard, the construct further comprises regulatory sequences, including, for example, a promoter, operably linked to the sequence. Large numbers of suitable vectors and promoters are known to those of skill

- 52 -

in the art, and there are many commercially available vectors suitable for use in the present invention.

The following vectors, which are commercially available, are provided by way of example. Among vectors preferred for use in bacteria are pQE70, pQE60 and pQE-9, available from Qiagen; pBS vectors, Phagescript vectors, Bluescript vectors, pNH8A, pNH16a, pNH18A, pNH46A, available from Stratagene; and ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 available from Pharmacia. Among preferred eukaryotic vectors are PWLNEO, pSV2CAT, pOG44, pXT1 and pSG available from Stratagene; and pSVK3, pBPV, pMSG and pSVL available from Pharmacia. These vectors are listed solely by way of illustration of the many commercially available and well known vectors that are available to those of skill in the art for use in accordance with this aspect of the present invention. It will be appreciated by those of skill in the art upon reading this disclosure that any other plasmid or vector suitable for introduction, maintenance, propagation and/or expression of a LSG polynucleotide or polypeptide of the invention in a host may be used in this aspect of the invention.

Promoter regions can be selected from any desired gene using vectors that contain a reporter transcription unit lacking a promoter region, such as a chloramphenicol acetyl transferase ("cat") transcription unit, downstream of a restriction site or sites for introducing a candidate promoter fragment; i.e., a fragment that may contain a promoter. As is well known, introduction into the vector of a promoter-containing fragment at the restriction site upstream of the cat gene engenders production of CAT activity detectable by standard CAT assays. Vectors suitable to this end are well known and readily available. Two such vectors are pKK232-8 and pCM7. Thus, promoters for expression of LSG polynucleotides of the present invention include, not only well known and readily

- 53 -

available promoters, but also promoters that readily may be obtained by the foregoing technique, using a reporter gene.

Among known bacterial promoters suitable for expression of polynucleotides and polypeptides in accordance with the present invention are the *E. coli* lacI and lacZ promoters, the T3 and T7 promoters, the gpt promoter, the lambda PR, PL promoters and the trp promoter. Among known eukaryotic promoters suitable in this regard are the CMV immediate early promoter, the HSV thymidine kinase promoter, the early and late SV40 promoters, the promoters of retroviral LTRs, such as those of the Rous sarcoma virus ("RSV"), and metallothionein promoters, such as the mouse metallothionein-I promoter.

Selection of appropriate vectors and promoters for expression in a host cell is a well known procedure and the requisite techniques for expression vector construction, introduction of the vector into the host and expression in the host are routine skills in the art.

The present invention also relates to host cells containing the above-described constructs. The host cell can be a higher eukaryotic cell, such as a mammalian cell, or a lower eukaryotic cell, such as a yeast cell. Alternatively, the host cell can be a prokaryotic cell, such as a bacterial cell.

Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE-dextran mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection or other methods. Such methods are described in many standard laboratory manuals, such as Davis et al. BASIC METHODS IN MOLECULAR BIOLOGY, (1986).

Constructs in host cells can be used in a conventional manner to produce the gene product encoded by the recombinant sequence. Alternatively, LSG polypeptides

- 54 -

of the invention can be synthetically produced by conventional peptide synthesizers.

Mature proteins can be expressed in mammalian cells, yeast, bacteria, or other cells under the control of appropriate promoters. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described by Sambrook et al. cited elsewhere herein.

Generally, recombinant expression vectors will include origins of replication, a promoter derived from a highly-expressed gene to direct transcription of a downstream structural sequence, and a selectable marker to permit isolation of vector containing cells after exposure to the vector. Among suitable promoters are those derived from the genes that encode glycolytic enzymes such as 3-phosphoglycerate kinase ("PGK"), a-factor, acid phosphatase, and heat shock proteins, among others. Selectable markers include the ampicillin resistance gene of *E. coli* and the *trp1* gene of *S. cerevisiae*.

Transcription of DNA encoding the LSG polypeptides of the present invention by higher eukaryotes may be increased by inserting an enhancer sequence into the vector. Enhancers are cis-acting elements of DNA, usually about from 10 to 300 base pairs (bp) that act to increase transcriptional activity of a promoter in a given host cell-type. Examples of enhancers include the SV40 enhancer, which is located on the late side of the replication origin at bp 100 to 270, the cytomegalovirus early promoter enhancer, the polyoma enhancer on the late side of the replication origin, and adenovirus enhancers.

A polynucleotide of the present invention, encoding a heterologous structural sequence of a LSG polypeptide of the present invention, generally will be inserted into the

- 55 -

vector using standard techniques so that it is operably linked to the promoter for expression. The polynucleotide will be positioned so that the transcription start site is located appropriately 5' to a ribosome binding site. The  
5 ribosome binding site will be 5' to the AUG that initiates translation of the polypeptide to be expressed. Generally, there will be no other open reading frames that begin with an initiation codon, usually AUG, lying between the ribosome binding site and the initiating AUG. Also,  
10 generally, there will be a translation stop codon at the end of the polypeptide and there will be a polyadenylation signal and a transcription termination signal appropriately disposed at the 3' end of the transcribed region.

Appropriate secretion signals may be incorporated  
15 into the expressed polypeptide for secretion of the translated protein into the lumen of the endoplasmic reticulum, into the periplasmic space or into the extracellular environment. The signals may be endogenous to the polypeptide or they may be heterologous signals.

20 The polypeptide may be expressed in a modified form, such as a fusion protein, and may include not only secretion signals but also additional heterologous functional regions. Thus, for instance, a region of additional amino acids, particularly charged amino acids,  
25 may be added to the N-terminus of the polypeptide to improve stability and persistence in the host cell during purification or during subsequent handling and storage. A region also may be added to the polypeptide to facilitate purification. Such regions may be removed prior to final  
30 preparation of the polypeptide. The addition of peptide moieties to polypeptides to engender secretion or excretion, to improve stability and to facilitate purification, among others, are familiar and routine techniques in the art.

- 56 -

Suitable prokaryotic hosts for propagation, maintenance or expression of LSG polynucleotides and polypeptides in accordance with the invention include *Escherichia coli*, *Bacillus subtilis* and *Salmonella typhimurium*. Various species of *Pseudomonas*, *Streptomyces*, and *Staphylococcus* are suitable hosts in this regard. Many other hosts also known to those of skill may also be employed in this regard.

As a representative, but non-limiting example, useful expression vectors for bacterial use can comprise a selectable marker and bacterial origin of replication derived from commercially available plasmids comprising genetic elements of the well known cloning vector pBR322. Such commercial vectors include, for example, pKK223-3 (Pharmacia Fine Chemicals, Uppsala, Sweden) and GEM1 (Promega Biotec, Madison, Wis., USA). These pBR322 "backbone" sections are combined with an appropriate promoter and the structural sequence to be expressed. Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, where the selected promoter is inducible it is induced by appropriate means (e.g., temperature shift or exposure to chemical inducer) and cells are cultured for an additional period. Cells typically then are harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification. Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents, such methods are well know to those skilled in the art.

Various mammalian cell culture systems can be employed for expression, as well. An exemplary mammalian expression systems is the COS-7 line of monkey kidney fibroblasts described in Gluzman et al., Cell 23: 175

- 57 -

(1981). Other mammalian cell lines capable of expressing a compatible vector include for example, the C127, 3T3, CHO, HeLa, human kidney 293 and BHK cell lines. Mammalian expression vectors comprise an origin of replication, a suitable promoter and enhancer, and any ribosome binding sites, polyadenylation sites, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking non-transcribed sequences that are necessary for expression. In certain preferred embodiments in this regard DNA sequences derived from the SV40 splice sites, and the SV40 polyadenylation sites are used for required non-transcribed genetic elements of these types.

LSG polypeptides can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Most preferably, high performance liquid chromatography ("HPLC") is employed for purification. Well known techniques for refolding proteins may be employed to regenerate active conformation when the polypeptide is denatured during isolation and or purification.

LSG polypeptides of the present invention include naturally purified products, products of chemical synthetic procedures, and products produced by recombinant techniques from a prokaryotic or eukaryotic host, including, for example, bacterial, yeast, higher plant, insect and mammalian cells. Depending upon the host employed in a recombinant production procedure, the LSG polypeptides of the present invention may be glycosylated or may be non-glycosylated. In addition, LSG polypeptides of the invention may also include an initial modified methionine

- 58 -

residue, in some cases as a result of host-mediated processes.

LSG polynucleotides and polypeptides may be used in accordance with the present invention for a variety of applications, particularly those that make use of the chemical and biological properties of the LSGs. Additional applications relate to diagnosis and to treatment of disorders of cells, tissues and organisms. These aspects of the invention are illustrated further by the following discussion.

#### *Polynucleotide assays*

As discussed in some detail *supra*, this invention is also related to the use of LSG polynucleotides to detect complementary polynucleotides such as, for example, as a diagnostic reagent. Detection of a mutated form of LSG associated with a dysfunction will provide a diagnostic tool that can add to or define a diagnosis of a disease or susceptibility to a disease which results from under-expression, over-expression or altered expression of a LSG, such as, for example, a susceptibility to inherited lung cancer.

Individuals carrying mutations in a human LSG gene may be detected at the DNA level by a variety of techniques. Nucleic acids for diagnosis may be obtained from a patient's cells, such as from blood, urine, saliva, tissue biopsy and autopsy material. The genomic DNA may be used directly for detection or may be amplified enzymatically using PCR prior to analysis (Saiki et al., *Nature*, 324: 163-166 (1986)). RNA or cDNA may also be used in a similar manner. As an example, PCR primers complementary to a LSG polynucleotide of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 37 or 38 can be used to identify and analyze LSG expression and mutations. For example, deletions and insertions can be detected by a change in



- 59 -

size of the amplified product in comparison to the normal genotype. Point mutations can be identified by hybridizing amplified DNA to radiolabeled LSG RNA or alternatively, radiolabeled LSG antisense DNA sequences. Perfectly  
5 matched sequences can be distinguished from mismatched duplexes by RNase A digestion or by differences in melting temperatures.

Sequence differences between a reference gene and genes having mutations also may be revealed by direct DNA  
10 sequencing. In addition, cloned DNA segments may be employed as probes to detect specific DNA segments. The sensitivity of such methods can be greatly enhanced by appropriate use of PCR or another amplification method. For example, a sequencing primer is used with double-  
15 stranded PCR product or a single-stranded template molecule generated by a modified PCR. The sequence determination is performed by conventional procedures with radiolabeled nucleotide or by automatic sequencing procedures with fluorescent-tags.

20 Genetic testing based on DNA sequence differences may be achieved by detection of alterations in electrophoretic mobility of DNA fragments in gels, with or without denaturing agents. Small sequence deletions and insertions can be visualized by high resolution gel electrophoresis.  
25 DNA fragments of different sequences may be distinguished on denaturing formamide gradient gels in which the mobilities of different DNA fragments are retarded in the gel at different positions according to their specific melting or partial melting temperatures (see, e.g., Myers  
30 et al., Science, 230: 1242 (1985)).

Sequence changes at specific locations also may be revealed by nuclease protection assays, such as RNase and S1 protection or the chemical cleavage method (e.g., Cotton et al., Proc. Natl. Acad. Sci., USA, 85: 4397-4401 (1985)).

- 60 -

Thus, the detection of a specific DNA sequence may be achieved by methods such as hybridization, RNase protection, chemical cleavage, direct DNA sequencing or the use of restriction enzymes, (e.g., restriction fragment length polymorphisms ("RFLP") and Southern blotting of genomic DNA. In addition to more conventional gel-electrophoresis and DNA sequencing, mutations also can be detected by *in situ* analysis.

#### *Chromosome assays*

10       The LSG sequences of the present invention are also valuable for chromosome identification. There is a need for identifying particular sites on the chromosome and few chromosome marking reagents based on actual sequence data (repeat polymorphisms) are presently available for marking  
15       chromosomal location. Each LSG sequence of the present invention is specifically targeted to and can hybridize with a particular location on an individual human chromosome. Thus, the LSGs can be used in the mapping of DNAs to chromosomes, an important first step in correlating  
20       sequences with genes associated with disease.

      In certain preferred embodiments in this regard, the cDNA herein disclosed is used to clone genomic DNA of a LSG of the present invention. This can be accomplished using a variety of well known techniques and libraries, which  
25       generally are available commercially. The genomic DNA is used for *in situ* chromosome mapping using well known techniques for this purpose.

      In some cases, sequences can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp) from the  
30       cDNA. Computer analysis of the 3' untranslated region of the gene is used to rapidly select primers that do not span more than one exon in the genomic DNA, thus complicating the amplification process. These primers are then used for PCR screening of somatic cell hybrids containing individual  
35       human chromosomes. Only those hybrids containing the human

- 61 -

gene corresponding to the primer will yield an amplified fragment.

PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular DNA to a particular  
5 chromosome. Using the present invention with the same oligonucleotide primers, sublocalization can be achieved with panels of fragments from specific chromosomes or pools of large genomic clones in an analogous manner. Other mapping strategies that can similarly be used to map to its  
10 chromosome include *in situ* hybridization, prescreening with labeled flow-sorted chromosomes and preselection by hybridization to construct chromosome specific-cDNA libraries.

Fluorescence *in situ* hybridization ("FISH") of a cDNA  
15 clone to a metaphase chromosomal spread can be used to provide a precise chromosomal location in one step. This technique can be used with cDNA as short as 50 or 60 bp. This technique is described by Verma et al. (HUMAN CHROMOSOMES: A MANUAL OF BASIC TECHNIQUES, Pergamon Press,  
20 New York (1988)).

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. Such data are found, for example, in V. McKusick, MENDELIAN  
25 INHERITANCE IN MAN, available on line through Johns Hopkins University, Welch Medical Library. The relationship between genes and diseases that have been mapped to the same chromosomal region are then identified through linkage analysis (coinheritance of physically adjacent genes).

30 Next, it is necessary to determine the differences in the cDNA or genomic sequence between affected and unaffected individuals. If a mutation is observed in some or all of the affected individuals but not in any normal individuals, then the mutation is likely to be the  
35 causative agent of the disease.

- 62 -

With current resolution of physical mapping and genetic mapping techniques, a cDNA precisely localized to a chromosomal region associated with the disease could be one of between 50 and 500 potential causative genes. (This  
5 assumes 1 megabase mapping resolution and one gene per 20 kb).

#### *Polypeptide assays*

As described in some detail *supra*, the present invention also relates to diagnostic assays such as  
10 quantitative and diagnostic assays for detecting levels of LSG polypeptide in cells and tissues, and biological fluids such as blood and urine, including determination of normal and abnormal levels. Thus, for instance, a diagnostic assay in accordance with the present invention for detecting  
15 over-expression or under-expression of a LSG polypeptide compared to normal control tissue samples may be used to detect the presence of neoplasia. Assay techniques that can be used to determine levels of a protein, such as a LSG polypeptide of the present invention, in a sample derived  
20 from a host are well-known to those of skill in the art. Such assay methods include radioimmunoassays, competitive-binding assays, Western Blot analysis and ELISA assays. Among these ELISAs frequently are preferred.

For example, antibody-sandwich ELISAs are used to  
25 detect polypeptides in a sample, preferably a biological sample. Wells of a microtiter plate are coated with specific antibodies, at a final concentration of 0.2 to 10  $\mu\text{g/ml}$ . The antibodies are either monoclonal or polyclonal and are produced by methods as described herein. The wells  
30 are blocked so that non-specific binding of the polypeptide to the well is reduced. The coated wells are then incubated for > 2 hours at room temperature with a sample containing the LSG polypeptide. Preferably, serial dilutions of the sample should be used to validate results.  
35 The plates are then washed three times with deionized or

- 63 -

distilled water to remove unbounded polypeptide. Next, 50  $\mu$ l of specific antibody-alkaline phosphatase conjugate, at a concentration of 25-400 ng, is added and incubated for 2 hours at room temperature. The plates are again washed  
5 three times with deionized or distilled water to remove unbounded conjugate. 4-methylumbelliferyl phosphate (MUP) or p-nitrophenyl phosphate (NPP) substrate solution (75  $\mu$ l) is then added to each well and the plate is incubated 1 hour at room temperature. The reaction is measured by a  
10 microtiter plate reader. A standard curve is prepared using serial dilutions of a control sample, and polypeptide concentration is plotted on the X-axis (log scale) while fluorescence or absorbance is plotted on the Y-axis (linear scale). The concentration of the LSG polypeptide in the  
15 sample is interpolated using the standard curve.

#### **Antibodies**

As discussed in some detail *supra*, LSG polypeptides, their fragments or other derivatives, or analogs thereof, or cells expressing them can be used as an immunogen to  
20 produce antibodies thereto. These antibodies can be polyclonal or monoclonal antibodies. The present invention also includes chimeric, single chain, and humanized antibodies, as well as Fab fragments, or the product of an Fab expression library. Various procedures known in the art  
25 may be used for the production of such antibodies and fragments.

A variety of methods for antibody production are set forth in Current Protocols, Chapter 2.

For example, cells expressing a LSG polypeptide of  
30 the present invention can be administered to an animal to induce the production of sera containing polyclonal antibodies. In a preferred method, a preparation of the secreted protein is prepared and purified to render it substantially free of natural contaminants. This  
35 preparation is then introduced into an animal in order to

- 64 -

produce polyclonal antisera of greater specific activity. The antibody obtained will bind with the LSG polypeptide itself. In this manner, even a sequence encoding only a fragment of the LSG polypeptide can be used to generate  
5 antibodies binding the whole native polypeptide. Such antibodies can then be used to isolate the LSG polypeptide from tissue expressing that LSG polypeptide.

Alternatively, monoclonal antibodies can be prepared. Examples of techniques for production of monoclonal  
10 antibodies include, but are not limited to, the hybridoma technique (Kohler, G. and Milstein, C., Nature 256: 495-497 (1975), the trioma technique, the human B-cell hybridoma technique (Kozbor et al., Immunology Today 4: 72 (1983) and (Cole et al., pg. 77-96 in MONOCLONAL ANTIBODIES AND CANCER  
15 THERAPY, Alan R. Liss, Inc. (1985)). The EBV-hybridoma technique is useful in production of human monoclonal antibodies.

Hybridoma technologies have also been described by Khler et al. (Eur. J. Immunol. 6: 511 (1976)) Khler et al.  
20 (Eur. J. Immunol. 6: 292 (1976)) and Hammerling et al. (in: Monoclonal Antibodies and T-Cell Hybridomas, Elsevier, N. Y., pp. 563-681 (1981)). In general, such procedures involve immunizing an animal (preferably a mouse) with LSG polypeptide or, more preferably, with a secreted LSG  
25 polypeptide-expressing cell. Such cells may be cultured in any suitable tissue culture medium; however, it is preferable to culture cells in Earle's modified Eagle's medium supplemented with 10% fetal bovine serum (inactivated at about 56°C), and supplemented with about 10  
30 g/l of nonessential amino acids, about 1,000 U/ml of penicillin, and about 100 µg/ml of streptomycin. The splenocytes of such mice are extracted and fused with a suitable myeloma cell line. Any suitable myeloma cell line may be employed in accordance with the present invention;  
35 however, it is preferable to employ the parent myeloma cell

- 65 -

line (SP20), available from the ATCC. After fusion, the resulting hybridoma cells are selectively maintained in HAT medium, and then cloned by limiting dilution as described by Wands et al. (Gastroenterology 80: 225-232 (1981)).

- 5 The hybridoma cells obtained through such a selection are then assayed to identify clones which secrete antibodies capable of binding the polypeptide.

Alternatively, additional antibodies capable of binding to the polypeptide can be produced in a two-step  
10 procedure using anti-idiotypic antibodies. Such a method makes use of the fact that antibodies are themselves antigens, and therefore, it is possible to obtain an antibody which binds to a second antibody. In accordance with this method, protein specific antibodies are used to  
15 immunize an animal, preferably a mouse. The splenocytes of such an animal are then used to produce hybridoma cells, and the hybridoma cells are screened to identify clones which produce an antibody whose ability to bind to the protein-specific antibody can be blocked by the  
20 polypeptide. Such antibodies comprise anti-idiotypic antibodies to the protein specific antibody and can be used to immunize an animal to induce formation of further protein-specific antibodies.

Techniques described for the production of single  
25 chain antibodies (U.S. Patent 4,946,778) can also be adapted to produce single chain antibodies to immunogenic polypeptide products of this invention. Also, transgenic mice, as well as other nonhuman transgenic animals, may be used to express humanized antibodies to immunogenic  
30 polypeptide products of this invention.

It will be appreciated that Fab, F(ab')<sub>2</sub> and other fragments of the antibodies of the present invention may also be used according to the methods disclosed herein. Such fragments are typically produced by proteolytic  
35 cleavage, using enzymes such as papain (to produce Fab

- 66 -

fragments) or pepsin (to produce F(ab')<sub>2</sub> fragments). Alternatively, secreted protein-binding fragments can be produced through the application of recombinant DNA technology or through synthetic chemistry.

- 5 For *in vivo* use of antibodies in humans, it may be preferable to use "humanized" chimeric monoclonal antibodies. Such antibodies can be produced using genetic constructs derived from hybridoma cells producing the monoclonal antibodies described above. Methods for
- 10 producing chimeric antibodies are known in the art (See, for review, Morrison, Science 229: 1202 (1985); Oi et al., BioTechniques 4: 214 (1986); Cabilly et al., U. S. Patent 4,816,567; Taniguchi et al., EP 171496; Morrison et al., EP 173494; Neuberger et al., WO 8601533; Robinson et al., WO
- 15 8702671; Boulianne et al., Nature 312: 643 (1984); Neuberger et al., Nature 314: 268 (1985).)

The above-described antibodies may be employed to isolate or to identify clones expressing LSG polypeptides or purify LSG polypeptides of the present invention by

20 attachment of the antibody to a solid support for isolation and/or purification by affinity chromatography. As discussed in more detail *supra*, antibodies specific against a LSG may also be used to image tumors, particularly cancer of the lung, in patients suffering from cancer. Such

25 antibodies may also be used therapeutically to target tumors expressing a LSG.

Preferred exemplary antigenic epitopes of LSGs of the present invention which have been identified are depicted below. The antigenicity index (AI avg) used is Jameson-

30 Wolf. In some embodiment, it may be preferred to raise antibodies against these regions of the LSGs.

LSG of SEQ ID NO:39

	positions	AI avg	length
	176-220	1.37	45
35	399-410	1.18	12
	301-317	1.13	17



- 67 -

	370-391	1.13	22
	23-34	1.07	12
	149-174	1.00	26
	51-67	1.00	17
5	LSG of SEQ ID NO:42		
	positions	AI avg	length
	453-465	1.25	13
	399-409	1.25	11
	572-584	1.20	13
10	874-887	1.18	14
	226-235	1.15	10
	30-51	1.09	22
	910-920	1.07	11
	991-1010	1.06	20
15	655-668	1.06	14
	362-373	1.00	12
	LSG of SEQ ID NO:44		
	positions	AI avg	length
	134-160	1.23	27
20	415-436	1.17	22
	485-515	1.16	31
	459-474	1.10	16
	200-210	1.08	11
	535-562	1.04	28
25	91-115	1.04	25
	523-532	1.02	10
	8-20	1.01	13
	LSG of SEQ ID NO:45		
	positions	AI avg	length
30	563-586	1.19	24
	395-408	1.09	14
	130-139	1.04	10
	117-127	1.02	11
	165-189	1.01	25
35	LSG of SEQ ID NO:46		
	positions	AI avg	length
	122-137	1.10	16
	LSG of SEQ ID NO:47		
	positions	AI avg	length
40	1045-1054	1.12	10
	845-880	1.10	36
	919-945	1.10	27
	1376-1418	1.10	43
	144-164	1.10	21
45	814-835	1.09	22
	706-755	1.06	50
	401-416	1.05	16
	445-491	1.04	47
	1061-1085	1.03	25
50	422-442	1.02	21
	LSG of SEQ ID NO:48		

- 68 -

	positions	AI avg	length
	340-362	1.05	23
	155-164	1.01	10
	228-240	1.00	13
5	3-14	1.00	12
	LSG of SEQ ID NO:49		
	positions	AI avg	length
	189-204	1.08	16
	LSG of SEQ ID NO:50		
10	positions	AI avg	length
	134-143	1.21	10
	23-45	1.01	23
	LSG of SEQ ID NO:51		
	positions	AI avg	length
15	53-68	1.14	16
	LSG of SEQ ID NO:53		
	positions	AI avg	length
	367-392	1.32	26
	491-504	1.07	14
20	14-35	1.04	22
	275-284	1.03	10
	208-219	1.03	12
	439-456	1.02	18
	LSG of SEQ ID NO:54		
25	positions	AI avg	length
	1671-1681	1.35	11
	453-465	1.26	13
	1748-1759	1.23	12
	1725-1738	1.19	14
30	1804-1825	1.15	22
	1644-1655	1.13	12
	1281-1295	1.12	15
	1532-1545	1.11	14
	1351-1369	1.07	19
35	1040-1062	1.06	23
	1334-1347	1.05	14
	145-155	1.05	11
	1121-1132	1.05	12
	1307-1318	1.02	12
40	1376-1408	1.02	33
	650-660	1.01	11
	802-823	1.00	22
	714-735	1.00	22
	1885-1898	1.00	14
45	1967-1976	1.00	10
	LSG of SEQ ID NO:55		
	positions	AI avg	length
	297-311	1.31	15
	328-344	1.25	17
50	16-25	1.20	10
	96-113	1.12	18

- 69 -

	381-393	1.12	13
	236-250	1.10	15
	354-364	1.09	11
	441-451	1.07	11
5	274-291	1.00	18
LSG of SEQ ID NO:56			
	positions	AI avg	length
	197-210	1.03	14
	318-328	1.02	11

10

*LSG binding molecules and assays*

This invention also provides a method for identification of molecules, such as receptor molecules, that bind LSGs. Genes encoding proteins that bind LSGs, such as receptor proteins, can be identified by numerous methods known to those of skill in the art. Examples include, but are not limited to, ligand panning and FACS sorting. Such methods are described in many laboratory manuals such as, for instance, Coligan et al., Current Protocols in Immunology 1(2): Chapter 5 (1991).

Expression cloning may also be employed for this purpose. To this end, polyadenylated RNA is prepared from a cell responsive to a LSG of the present invention. A cDNA library is created from this RNA and the library is divided into pools. The pools are then transfected individually into cells that are not responsive to a LSG of the present invention. The transfected cells then are exposed to labeled LSG. LSG polypeptides can be labeled by a variety of well-known techniques including, but not limited to, standard methods of radio-iodination or inclusion of a recognition site for a site-specific protein kinase. Following exposure, the cells are fixed and binding of labeled LSG is determined. These procedures conveniently are carried out on glass slides. Pools containing labeled LSG are identified as containing cDNA that produced LSG-binding cells. Sub-pools are then prepared from these positives, transfected into host cells and screened as described above. Using an iterative sub-pooling and re-

- 70 -

screening process, one or more single clones that encode the putative binding molecule, such as a receptor molecule, can be isolated.

Alternatively a labeled ligand can be photoaffinity  
5 linked to a cell extract, such as a membrane or a membrane  
extract, prepared from cells that express a molecule that  
it binds, such as a receptor molecule. Cross-linked  
material is resolved by polyacrylamide gel electrophoresis  
("PAGE") and exposed to X-ray film. The labeled complex  
10 containing the ligand-receptor can be excised, resolved  
into peptide fragments, and subjected to protein  
microsequencing. The amino acid sequence obtained from  
microsequencing can be used to design unique or degenerate  
oligonucleotide probes to screen cDNA libraries to identify  
15 genes encoding the putative receptor molecule.

Polypeptides of the invention also can be used to  
assess LSG binding capacity of LSG binding molecules, such  
as receptor molecules, in cells or in cell-free  
preparations.

#### 20 *Agonists and antagonists - assays and molecules*

The invention also provides a method of screening  
compounds to identify those which enhance or block the  
action of a LSG on cells. By "compound", as used herein,  
it is meant to be inclusive of small organic molecules,  
25 peptides, polypeptides and antibodies as well as any other  
candidate molecules which have the potential to enhance or  
agonize or block or antagonize the action of LSG on cells.  
As used herein, an agonist is a compound which increases  
the natural biological functions of a LSG or which  
30 functions in a manner similar to a LSG, while an  
antagonist, as used herein, is a compound which decreases  
or eliminates such functions. Various known methods for  
screening for agonists and/or antagonists can be adapted  
for use in identifying LSG agonist or antagonists.

- 71 -

For example, a cellular compartment, such as a membrane or a preparation thereof, such as a membrane-preparation, may be prepared from a cell that expresses a molecule that binds a LSG, such as a molecule of a signaling or regulatory pathway modulated by LSG. The preparation is incubated with labeled LSG in the absence or the presence of a compound which may be a LSG agonist or antagonist. The ability of the compound to bind the binding molecule is reflected in decreased binding of the labeled ligand. Compounds which bind gratuitously, i.e., without inducing the effects of a LSG upon binding to the LSG binding molecule are most likely to be good antagonists. Compounds that bind well and elicit effects that are the same as or closely related to LSG are agonists. LSG-like effects of potential agonists and antagonists may be measured, for instance, by determining activity of a second messenger system following interaction of the candidate molecule with a cell or appropriate cell preparation, and comparing the effect with that of LSG or molecules that elicit the same effects as LSG. Second messenger systems that may be useful in this regard include, but are not limited to, AMP guanylate cyclase, ion channel or phosphoinositide hydrolysis second messenger systems.

Another example of an assay for LSG antagonists is a competitive assay that combines LSG and a potential antagonist with membrane-bound LSG receptor molecules or recombinant LSG receptor molecules under appropriate conditions for a competitive inhibition assay. LSG can be labeled, such as by radioactivity, such that the number of LSG molecules bound to a receptor molecule can be determined accurately to assess the effectiveness of the potential antagonist.

Potential antagonists include small organic molecules, peptides, polypeptides and antibodies that bind

- 72 -

to a LSG polypeptide of the invention and thereby inhibit or extinguish its activity. Potential antagonists also may be small organic molecules, a peptide, a polypeptide such as a closely related protein or antibody that binds the  
5 same sites on a binding molecule, such as a receptor molecule, without inducing LSG-induced activities, thereby preventing the action of LSG by excluding LSG from binding.

Potential antagonists include small molecules which bind to and occupy the binding site of the LSG polypeptide  
10 thereby preventing binding to cellular binding molecules, such as receptor molecules, such that normal biological activity is prevented. Examples of small molecules include but are not limited to small organic molecules, peptides or peptide-like molecules.

15 Other potential antagonists include antisense molecules. Antisense technology can be used to control gene expression through antisense DNA or RNA or through triple-helix formation. Antisense techniques are discussed, for example, in Okano, J. Neurochem. 56: 560 (1991);  
20 OLIGODEOXYNUCLEOTIDES AS ANTISENSE INHIBITORS OF GENE EXPRESSION, CRC Press, Boca Raton, Fla. (1988). Triple helix formation is discussed in, for instance Lee et al., Nucleic Acids Research 6: 3073 (1979); Cooney et al., Science 241: 456 (1988); and Dervan et al., Science 251:  
25 1360 (1991). The methods are based on binding of a polynucleotide to a complementary DNA or RNA. For example, the 5' coding portion of a polynucleotide that encodes a mature LSG polypeptide of the present invention may be used to design an antisense RNA oligonucleotide of from about 10  
30 to 40 base pairs in length. A DNA oligonucleotide is designed to be complementary to a region of the gene involved in transcription thereby preventing transcription and the production of a LSG polypeptide. The antisense RNA oligonucleotide hybridizes to the mRNA *in vivo* and blocks  
35 translation of the mRNA molecule into a LSG polypeptide.

- 73 -

The oligonucleotides described above can also be delivered to cells such that the antisense RNA or DNA may be expressed *in vivo* to inhibit production of a LSG.

### *Compositions*

5       The present invention also relates to compositions comprising a LSG polynucleotide or a LSG polypeptide or an agonist or antagonist thereof.

For example, a LSG polynucleotide, polypeptide or an agonist or antagonist thereof of the present invention may  
10 be employed in combination with a non-sterile or sterile carrier or carriers for use with cells, tissues or organisms, such as a pharmaceutical carrier suitable for administration to a subject. Such compositions comprise, for instance, a media additive or a therapeutically  
15 effective amount of a polypeptide of the invention and a pharmaceutically acceptable carrier or excipient. Such carriers may include, but are not limited to, saline, buffered saline, dextrose, water, glycerol, ethanol and combinations thereof. The formulation should suit the mode  
20 of administration.

Compositions of the present invention will be formulated and dosed in a fashion consistent with good medical practice, taking into account the clinical condition of the individual patient (especially the side  
25 effects of treatment with the polypeptide or other compound alone), the site of delivery, the method of administration, the scheduling of administration, and other factors known to practitioners. The "effective amount" for purposes herein is thus determined by such considerations.

30       As a general proposition, the total pharmaceutically effective amount of secreted polypeptide administered parenterally per dose will be in the range of about 1,  $\mu\text{g/kg/day}$  to 10  $\text{mg/kg/day}$  of patient body weight, although, as noted above, this will be subject to therapeutic  
35 discretion. More preferably, this dose is at least 0.01

- 74 -

mg/kg/day, and most preferably for humans between about 0.01 and 1 mg/kg/day for the hormone. If given continuously, the polypeptide or other compound is typically administered at a dose rate of about 1  $\mu$ g/kg/hour to about 50 mg/kg/hour, either by 1-4 injections per day or by continuous subcutaneous infusion, for example, using a mini-pump. An intravenous bag solution may also be employed. The length of treatment needed to observe changes and the interval following treatment for responses to occur appears to vary depending on the desired effect.

Pharmaceutical compositions containing the secreted protein of the invention are administered orally, rectally, parenterally, intracistemally, intravaginally, intraperitoneally, topically (as by powders, ointments, gels, drops or transdermal patch), buccally, or as an oral or nasal spray. "Pharmaceutically acceptable carrier" refers to a non-toxic solid, semisolid or liquid filler, diluent, encapsulating material or formulation auxiliary of any type. The term "parenteral" as used herein refers to modes of administration which include intravenous, intramuscular, intraperitoneal, intrasternal, subcutaneous and intraarticular injection and infusion.

The polypeptide or other compound is also suitably administered by sustained-release systems. Suitable examples of sustained-release compositions include semipermeable polymer matrices in the form of shaped articles, e. g., films, or microcapsules. Sustained-release matrices include polylactides (U.S. Patent 3,773,919 and EP 58481), copolymers of L-glutamic acid and gamma-ethyl-L-glutamate (Sidman, U. et al., Biopolymers 22: 547-556 (1983)), poly (2-hydroxyethyl methacrylate) (R. Langer et al., J. Biomed. Mater. Res. 15: 167-277 (1981), and R. Langer, Chem. Tech. 12: 98-105 (1982)), ethylene vinyl acetate (R. Langer et al.) and poly-D- (-)-3-hydroxybutyric acid (EP 133,988). Sustained-release compositions also



- 75 -

include liposomally entrapped polypeptides. Liposomes containing the polypeptide or other compound are prepared by well known methods (Epstein et al., Proc. Natl. Acad. Sci. USA 82: 3688-3692 (1985); Hwang et al., Proc. Natl. Acad. Sci. USA 77: 4030-4034 (1980); EP 52322; EP 36676; EP 5 88046; EP 143949; EP 142641; Japanese Pat. Appl. 83-118008; U.S. Patent 4,485,045 and 4,544,545; and EP 102324). Ordinarily, the liposomes are of the small (about 200-800 Angstroms) unilamellar type in which the lipid content is 10 greater than about 30 mol. percent cholesterol, the selected proportion being adjusted for the optimal therapy.

For parenteral administration, in one embodiment, the polypeptide or other compound is formulated generally by mixing it at the desired degree of purity, in a unit dosage 15 injectable form (solution, suspension, or emulsion), with a pharmaceutically acceptable carrier, i.e., one that is non-toxic to recipients at the dosages and concentrations employed and is compatible with other ingredients of the formulation.

20 For example, the formulation preferably does not include oxidizing agents and other compounds that are known to be deleterious to the polypeptide or other compound.

Generally, the formulations are prepared by contacting the polypeptide or other compound uniformly and 25 intimately with liquid carriers or finely divided solid carriers or both. Then, if necessary, the product is shaped into the desired formulation. Preferably the carrier is a parenteral carrier, more preferably a solution that is isotonic with the blood of the recipient. Examples of such 30 carrier vehicles include water, saline, Ringer's solution, and dextrose solution. Non-aqueous vehicles such as fixed oils and ethyl oleate are also useful herein, as well as liposomes.

The carrier suitably contains minor amounts of 35 additives such as substances that enhance isotonicity and

- 76 -

chemical stability. Such materials are non-toxic to recipients at the dosages and concentrations employed, and include buffers such as phosphate, citrate, succinate, acetic acid, and other organic acids or their salts; 5 antioxidants such as ascorbic acid; low molecular weight (less than about ten residues) polypeptides, e. g., polyarginine or tripeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids, such as glycine, 10 glutamic acid, aspartic acid, or arginine; monosaccharides, disaccharides, and other carbohydrates including cellulose or its derivatives, glucose, mannose, or dextrans; chelating agents such as EDTA; sugar alcohols such as mannitol or sorbitol; counterions such as sodium; and/or 15 nonionic surfactants such as polysorbates, poloxamers, or PEG.

The polypeptide or other compound is typically formulated in such vehicles at a concentration of about 0.1 mg/ml to 100 mg/ml, preferably 1-10 mg/ml, at a pH of about 20 3 to 8. It will be understood that the use of certain of the foregoing excipients, carriers, or stabilizers will result in the formation of polypeptide salts or salts of the other compounds.

Any polypeptide to be used for therapeutic 25 administration should be sterile. Sterility is readily accomplished by filtration through sterile filtration membranes (e. g., 0.2 micron membranes). Therapeutic polypeptide compositions generally are placed into a container having a sterile access port, for example, an 30 intravenous solution bag or vial having a stopper pierceable by a hypodermic injection needle.

Polypeptides ordinarily will be stored in unit or multi-dose containers, for example, sealed ampules or vials, as an aqueous solution or as a lyophilized 35 formulation for reconstitution. As an example of a

- 77 -

lyophilized formulation, 10-ml vials are filled with 5 ml of sterile-filtered 1 % (w/v) aqueous polypeptide solution, and the resulting mixture is lyophilized. The infusion solution is prepared by reconstituting the lyophilized polypeptide using bacteriostatic Water-for-Injection.

#### Kits

The invention further relates to pharmaceutical packs and kits comprising one or more containers filled with one or more of the ingredients of the aforementioned compositions of the invention. Associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, reflecting approval by the agency of the manufacture, use or sale of the product for human administration.

#### Administration

LSG polypeptides or polynucleotides or other compounds, preferably agonists or antagonists thereof of the present invention may be employed alone or in conjunction with other compounds, such as therapeutic compounds.

The pharmaceutical compositions may be administered in any effective, convenient manner including, for instance, administration by topical, oral, anal, vaginal, intravenous, intraperitoneal, intramuscular, subcutaneous, intranasal or intradermal routes among others.

The pharmaceutical compositions generally are administered in an amount effective for treatment or prophylaxis of a specific indication or indications. In general, the compositions are administered in an amount of at least about 10  $\mu\text{g/kg}$  body weight. However, it will be appreciated that optimum dosage will be determined by standard methods for each treatment modality and indication, taking into account the indication, its

- 78 -

severity, route of administration, complicating conditions and the like.

It will be appreciated that conditions caused by a decrease in the standard or normal expression level of a LSG polypeptide in an individual can be treated by administering the LSG polypeptide of the present invention, preferably in the secreted form, or an agonist thereof. Thus, the invention also provides a method of treatment of an individual in need of an increased level of a LSG polypeptide comprising administering to such an individual a pharmaceutical composition comprising an amount of the LSG polypeptide or an agonist thereof to increase the activity level of the LSG polypeptide in such an individual. For example, a patient with decreased levels of a LSG polypeptide may receive a daily dose 0.1-100  $\mu\text{g/kg}$  of a LSG polypeptide or agonist thereof for six consecutive days. Preferably, if a LSG polypeptide is administered it is in the secreted form.

Compositions of the present invention can also be administered to treating increased levels of a LSG polypeptide. For example, antisense technology can be used to inhibit production of a LSG polypeptide of the present invention. This technology is one example of a method of decreasing levels of a polypeptide, preferably a secreted form, due to a variety of etiologies, such as cancer. A patient diagnosed with abnormally increased levels of a polypeptide can be administered intravenously antisense polynucleotides at 0.5, 1.0, 1.5, 2.0 and 3.0  $\text{mg/kg}$  day for 21 days. This treatment is preferably repeated after a 7-day rest period if the treatment was well tolerated. Compositions comprising an antagonist of a LSG polypeptide can also be administered to decrease levels of LSG in a patient.

*Gene therapy*

- 79 -

The LSG polynucleotides, polypeptides, agonists and antagonists that are polypeptides may be employed in accordance with the present invention by expression of such polypeptides *in vivo*, in treatment modalities often  
5 referred to as "gene therapy."

Thus, for example, cells from a patient may be engineered with a polynucleotide, such as a DNA or RNA, encoding a polypeptide *ex vivo*, and the engineered cells then can be provided to a patient to be treated with the  
10 polypeptide. For example, cells may be engineered *ex vivo* by the use of a retroviral plasmid vector containing RNA encoding a polypeptide of the present invention. Such methods are well-known in the art and their use in the present invention will be apparent from the teachings  
15 herein.

Similarly, cells may be engineered *in vivo* for expression of a polypeptide *in vivo* by procedures known in the art. For example, a polynucleotide of the invention may be engineered for expression in a replication defective  
20 retroviral vector, as discussed *supra*. The retroviral expression construct then may be isolated and introduced into a packaging cell transduced with a retroviral plasmid vector containing RNA encoding a polypeptide of the present invention such that the packaging cell now produces  
25 infectious viral particles containing the gene of interest. These producer cells may be administered to a patient for engineering cells *in vivo* and expression of the polypeptide *in vivo*. These and other methods for administering a polypeptide of the present invention would be apparent to  
30 those skilled in the art upon reading the instant application.

Retroviruses from which the retroviral plasmid vectors herein above mentioned may be derived include, but are not limited to, Moloney Murine Leukemia Virus, spleen  
35 necrosis virus, retroviruses such as Rous Sarcoma Virus,

- 80 -

Harvey Sarcoma Virus, avian leukosis virus, gibbon ape leukemia virus, human immunodeficiency virus, adenovirus, Myeloproliferative Sarcoma Virus, and mammary tumor virus. In one embodiment, the retroviral plasmid vector is derived  
5 from Moloney Murine Leukemia Virus.

Such vectors will include one or more promoters for expressing the polypeptide. The selection of a suitable promoter will be apparent to those skilled in the art from the teachings contained herein. However, examples of  
10 suitable promoters which may be employed include, but are not limited to, the retroviral LTR, the SV40 promoter, the human cytomegalovirus (CMV) promoter described in Miller et al., Biotechniques 7: 980-990 (1989), and eukaryotic cellular promoters such as the histone, RNA polymerase III,  
15 and beta-actin promoters. Other viral promoters which may be employed include, but are not limited to, adenovirus promoters, thymidine kinase (TK) promoters, and B19 parvovirus promoters. Additional promoters which may be used include respiratory syncytial virus (RSV) promoter,  
20 inducible promoters such as the MMT promoter, the metallothionein promoter, heat shock promoters, the albumin promoter, the ApoA1 promoter, human globin promoters, viral thymidine kinase promoters such as the Herpes Simplex thymidine kinase promoter, retroviral LTRs, the beta-actin  
25 promoter, and human growth hormone promoters. The promoter also may be the native promoter which controls the gene encoding the polypeptide.

The nucleic acid sequence encoding the polypeptide of the present invention will be placed under the control of a  
30 suitable promoter.

In one embodiment, the retroviral plasmid vector is employed to transduce packaging cell lines to form producer cell lines. Examples of packaging cells which may be transfected include, but are not limited to, the PE501,  
35 PA317, Y-2, Y-AM, PA12, T19-14X, VT-19-17-H2, YCRE, YCRIP,

- 81 -

GP+E-86, GP+envAml2, and DAN cell lines as described in Miller, A., Human Gene Therapy 1: 5-14 (1990). The vector may be transduced into the packaging cells through any means known in the art. Such means include, but are not  
5 limited to, electroporation, the use of liposomes, and  $\text{CaPO}_4$  precipitation. Alternatively, the retroviral plasmid vector may be encapsulated into a liposome, or coupled to a lipid, and then administered to a host. The producer cell line will generate infectious retroviral vector particles  
10 which are inclusive of the nucleic acid sequence(s) encoding the polypeptides. Such retroviral vector particles then may be employed to transduce eukaryotic cells, either *in vitro* or *in vivo*. The transduced eukaryotic cells will express the nucleic acid sequence(s)  
15 encoding the polypeptide. Eukaryotic cells which may be transduced include, but are not limited to, embryonic stem cells, embryonic carcinoma cells, as well as hematopoietic stem cells, hepatocytes, fibroblasts, myoblasts, keratinocytes, endothelial cells, and bronchial epithelial  
20 cells.

An exemplary method of gene therapy involves transplantation of fibroblasts which are capable of expressing a LSG polypeptide or an agonist or antagonist thereof onto a patient. Generally fibroblasts are obtained  
25 from a subject by skin biopsy. The resulting tissue is placed in tissue-culture medium and separated into small pieces. Small chunks of the tissue are placed on a wet surface of a tissue culture flask, approximately ten pieces are placed in each flask. The flask is turned upside down,  
30 closed tight and left at room temperature over night. After 24 hours at room temperature, the flask is inverted and the chunks of tissue remain fixed to the bottom of the flask and fresh media (e. g., Ham's F12 media, with 10% FBS, penicillin and streptomycin) is added. The flasks are  
35 then incubated at 37°C for approximately one week. At this

- 82 -

time, fresh media is added and subsequently changed every several days. After an additional two weeks in culture, a monolayer of fibroblasts emerge. The monolayer is trypsinized and scaled into larger flasks. pMV-7

5 (Kirschmeier, P. T. et al., DNA, 7: 219-25 (1988)), flanked by the long terminal repeats of the Moloney murine sarcoma virus, is digested with EcoRI and HindIII and subsequently treated with calf intestinal phosphatase. The linear

10 vector is fractionated on agarose gel and purified, using glass beads. The cDNA encoding a LSG polypeptide of the present invention or an agonist or antagonist thereof can be amplified using PCR primers which correspond to their 5' and 3' end sequences respectively. Preferably, the 5' primer contains an EcoRI site and the 3' primer includes a

15 HindIII site. Equal quantities of the Moloney murine sarcoma virus linear backbone and the amplified EcoRI and HindIII fragment are added together in the presence of T4 DNA ligase. The resulting mixture is maintained under conditions appropriate for ligation of the two fragments.

20 The ligation mixture is then used to transform bacteria HB 101, which are then plated onto agar containing kanamycin for the purpose of confirming that the vector has the gene of interest properly inserted. Amphotropic pA317 or GP+aml2 packaging cells are grown in tissue culture to

25 confluent density in Dulbecco's Modified Eagles Medium (DMEM) with 10% calf serum (CS), penicillin and streptomycin. The MSV vector containing the gene is then added to the media and the packaging cells transduced with the vector. The packaging cells now produce infectious

30 viral particles containing the gene (the packaging cells are now referred to as producer cells). Fresh media is added to the transduced producer cells, and subsequently, the media is harvested from a 10 cm plate of confluent producer cells. The spent media, containing the infectious

35 viral particles, is filtered through a millipore filter to



- 83 -

remove detached producer cells and this media is then used to infect fibroblast cells. Media is removed from a sub-confluent plate of fibroblasts and quickly replaced with the media from the producer cells. This media is removed  
5 and replaced with fresh media. If the titer of virus is high, then virtually all fibroblasts will be infected and no selection is required. If the titer is very low, then it is necessary to use a retroviral vector that has a selectable marker, such as neo or his. Once the  
10 fibroblasts have been efficiently infected, the fibroblasts are analyzed to determine whether protein is produced. The engineered fibroblasts are then transplanted onto the host, either alone or after having been grown to confluence on cytodex 3 microcarrier beads.

15 Alternatively, in vivo gene therapy methods can be used to treat LSG related disorders, diseases and conditions. Gene therapy methods relate to the introduction of naked nucleic acid (DNA, RNA, and antisense DNA or RNA) sequences into an animal to increase or  
20 decrease the expression of the polypeptide.

For example, a LSG polynucleotide of the present invention or a nucleic acid sequence encoding an agonist or antagonist thereto may be operatively linked to a promoter or any other genetic elements necessary for the expression  
25 of the polypeptide by the target tissue. Such gene therapy and delivery techniques and methods are known in the art, see, for example, WO 90/11092, WO 98/11779; U.S. Patents 5,693,622, 5,705,151, and 5,580,859; Tabata H. et al. (1997) Cardiovasc. Res. 35 (3): 470-479, Chao J et al.  
30 (1997) Pharmacol. Res. 35 (6): 517-522, Wolff J. A. (1997) Neuromuscul. Disord. 7 (5): 314-318, Schwartz B. et al. (1996) Gene Ther. 3 (5): 405-411, Tsurumi Y. et al. (1996) Circulation 94 (12): 3281-3290 (incorporated herein by reference). The polynucleotide constructs may be delivered  
35 by any method that delivers injectable materials to the

- 84 -

cells of an animal, such as, injection into the interstitial space of tissues (heart, muscle, skin, lung, liver, intestine and the like). The polynucleotide constructs can be delivered in a pharmaceutically acceptable liquid or aqueous carrier.

The term "naked" polynucleotide, DNA or RNA, refers to sequences that are free from any delivery vehicle that acts to assist, promote, or facilitate entry into the cell, including viral sequences, viral particles, liposome formulations, lipofectin or precipitating agents and the like. However, polynucleotides may also be delivered in liposome formulations (such as those taught in Felgner P. L. et al. (1995) Ann. NY Acad. Sci. 772: 126-139 and Abdallah B. et al. (1995) Biol. Cell 85 (1): 1-7) which can be prepared by methods well known to those skilled in the art.

The polynucleotide vector constructs used in the gene therapy method are preferably constructs that will not integrate into the host genome nor will they contain sequences that allow for replication. Any strong promoter known to those skilled in the art can be used for driving the expression of DNA. Unlike other gene therapies techniques, one major advantage of introducing naked nucleic acid sequences into target cells is the transitory nature of the polynucleotide synthesis in the cells. Studies have shown that non-replicating DNA sequences can be introduced into cells to provide production of the desired polypeptide for periods of up to six months.

The polynucleotide construct can be delivered to the interstitial space of tissues within the an animal, including of muscle, skin, brain, lung, liver, spleen, bone marrow, thymus, heart, lymph, blood, bone, cartilage, pancreas, kidney, gall bladder, stomach, intestine, testis, ovary, uterus, rectum, nervous system, eye, gland, and connective tissue. Interstitial space of the tissues

- 85 -

comprises the intercellular fluid, mucopolysaccharide matrix among the reticular fibers of organ tissues, elastic fibers in the walls of vessels or chambers, collagen fibers of fibrous tissues, or that same matrix within connective tissue ensheathing muscle cells or in the lacunae of bone. It is similarly the space occupied by the plasma of the circulation and the lymph fluid of the lymphatic channels. Delivery to the interstitial space of muscle tissue is preferred. The polynucleotide construct may be conveniently delivered by injection into the tissues comprising these cells. They are preferably delivered to and expressed in persistent, non-dividing cells which are differentiated, although delivery and expression may be achieved in non-differentiated or less completely differentiated cells, such as, for example, stem cells of blood or skin fibroblasts. *In vivo* muscle cells are particularly competent in their ability to take up and express polynucleotides.

For the naked polynucleotide injection, an effective dosage amount of DNA or RNA will be in the range of from about 0.05  $\mu\text{g/kg}$  body weight to about 50 mg/kg body weight. Preferably the dosage will be from about 0.005 mg/kg to about 20 mg/kg and more preferably from about 0.05 mg/kg to about 5 mg/kg. Of course, as the artisan of ordinary skill will appreciate, this dosage will vary according to the tissue site of injection. The appropriate and effective dosage of nucleic acid sequence can readily be determined by those of ordinary skill in the art and may depend on the condition being treated and the route of administration. The preferred route of administration is by the parenteral route of injection into the interstitial space of tissues. However, other parenteral routes may also be used, such as, inhalation of an aerosol formulation particularly for delivery to lungs or bronchial tissues, throat or mucous membranes of the nose. In addition, naked polynucleotide

- 86 -

constructs can be delivered to arteries during angioplasty by the catheter used in the procedure.

The dose response effects of injected polynucleotide in muscle *in vivo* is determined as follows. Suitable  
5 template DNA for production of mRNA coding for polypeptide of the present invention is prepared in accordance with a standard recombinant DNA methodology. The template DNA, which may be either circular or linear, is either used as naked DNA or complexed with liposomes. The quadriceps  
10 muscles of mice are then injected with various amounts of the template DNA.

Five to six week old female and male Balb/C mice are anesthetized by intraperitoneal injection with 0.3 ml of 2.5% Avertin. A 1.5 cm incision is made on the anterior  
15 thigh, and the quadriceps muscle is directly visualized. The template DNA is injected in 0.1 ml of carrier in a 1 cc syringe through a 27 gauge needle over one minute, approximately 0.5 cm from the distal insertion site of the muscle into the knee and about 0.2 cm deep. A suture is  
20 placed over the injection site for future localization, and the skin is closed with stainless steel clips.

After an appropriate incubation time (e.g., 7 days) muscle extracts are prepared by excising the entire quadriceps. Every fifth 15  $\mu$ m cross-section of the  
25 individual quadriceps muscles is histochemically stained for protein expression. A time course for protein expression may be done in a similar fashion except that quadriceps from different mice are harvested at different times. Persistence of DNA in muscle following injection  
30 may be determined by Southern blot analysis after preparing total cellular DNA and HIRT supernatants from injected and control mice.

The results of the above experimentation in mice can be use to extrapolate proper dosages and other treatment  
35 parameters in humans and other animals using naked DNA.

- 87 -

*Nonhuman Transgenic Animals*

The LSG polypeptides of the invention can also be expressed in nonhuman transgenic animals. Nonhuman animals of any species, including, but not limited to, mice, rats, rabbits, hamsters, guinea pigs, pigs, micro-pigs, goats, sheep, cows and non-human primates, e. g., baboons, monkeys, and chimpanzees, may be used to generate transgenic animals. Any technique known in the art may be used to introduce the transgene (I. e., polynucleotides of the invention) into animals to produce the founder lines of transgenic animals. Such techniques include, but are not limited to, pronuclear microinjection (Paterson et al., Appl. Microbiol. Biotechnol. 40: 691-698 (1994); Carver et al., Biotechnology (NY) 11: 1263-1270 (1993); Wright et al., Biotechnology (NY) 9: 830-834 (1991); and Hoppe et al., U.S. Patent 4,873,191); retrovirus mediated gene transfer into germ lines (Van der Putten et al., Proc. Natl. Acad. Sci., USA 82: 6148-6152 (1985)), blastocysts or embryos; gene targeting in embryonic stem cells (Thompson et al., Cell 56: 313-321 (1989)); electroporation of cells or embryos (Lo, 1983, Mol. Cell. Biol. 3: 1803-1814 (1983)); introduction of the polynucleotides of the invention using a gene gun (see, e.g., Ulmer et al., Science 259: 1745 (1993); introducing nucleic acid constructs into embryonic pluripotent stem cells and transferring the stem cells back into the blastocyst; and sperm mediated gene transfer (Lavitrano et al., Cell 57: 717-723 (1989)). For a review of such techniques, see Gordon, "Transgenic Animals," Intl. Rev. Cytol. 115: 171-229 (1989), which is incorporated by reference herein in its entirety.

Any technique known in the art may be used to produce transgenic clones containing polynucleotides of the invention, for example, nuclear transfer into enucleated oocytes of nuclei from cultured embryonic, fetal, or adult

- 88 -

cells induced to quiescence (Campell et al., Nature 380: 64-66 (1996); Wilmut et al., Nature 385: 810813 (1997)).

The present invention provides for transgenic animals that carry the transgene in all their cells, as well as  
5 animals which carry the transgene in some, but not all their cells, i.e., mosaic or chimeric animals. The transgene may be integrated as a single transgene or as multiple copies such as in concatamers, e. g., head-to-head tandems or head-to-tail tandems. The transgene may also be  
10 selectively introduced into and activated in a particular cell type by following, for example, the teaching of Lasko et al. (Lasko et al., Proc. Natl. Acad. Sci. USA 89: 6232-6236 (1992)). The regulatory sequences required for such a cell-type specific activation will depend upon the  
15 particular cell type of interest, and will be apparent to those of skill in the art. When it is desired that the polynucleotide transgene be integrated into the chromosomal site of the endogenous gene, gene targeting is preferred. Briefly, when such a technique is to be utilized, vectors  
20 containing some nucleotide sequences homologous to the endogenous gene are designed for the purpose of integrating, via homologous recombination with chromosomal sequences, into and disrupting the function of the nucleotide sequence of the endogenous gene. The transgene  
25 may also be selectively introduced into a particular cell type, thus inactivating the endogenous gene in only that cell type, by following, for example, the teaching of Gu et al. (Science 265: 103-106 (1994)). The regulatory sequences required for such a cell-type specific  
30 inactivation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art.

Once transgenic animals have been generated, the expression of the recombinant gene may be assayed utilizing  
35 standard techniques. Initial screening may be accomplished

- 89 -

by Southern blot analysis or PCR techniques to analyze animal tissues to verify that integration of the transgene has taken place. The level of mRNA expression of the transgene in the tissues of the transgenic animals may also  
5 be assessed using techniques which include, but are not limited to, Northern blot analysis of tissue samples obtained from the animal, *in situ* hybridization analysis, and reverse transcriptase-PCR (rt-PCR). Samples of transgenic gene-expressing tissue may also be evaluated  
10 immunocytochemically or immunohistochemically using antibodies specific for the transgene product.

Once the founder animals are produced, they may be bred, inbred, outbred, or crossbred to produce colonies of the particular animal. Examples of such breeding  
15 strategies include, but are not limited to: outbreeding of founder animals with more than one integration site in order to establish separate lines; inbreeding of separate lines in order to produce compound transgenics that express the transgene at higher levels because of the effects of  
20 additive expression of each transgene; crossing of heterozygous transgenic animals to produce animals homozygous for a given integration site in order to both augment expression and eliminate the need for screening of animals by DNA analysis; crossing of separate homozygous  
25 lines to produce compound heterozygous or homozygous lines; and breeding to place the transgene on a distinct background that is appropriate for an experimental model of interest.

Transgenic animals of the invention have uses which  
30 include, but are not limited to, animal model systems useful in elaborating the biological function of LSG polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression of LSGs, and in screening for compounds effective in

- 90 -

ameliorating such LSG associated conditions and/or disorders.

#### ***Knock-Out Animals***

Endogenous gene expression can also be reduced by  
5 inactivating or "knocking out" the gene and/or its promoter  
using targeted homologous recombination (e. g., see  
Smithies et al., Nature 317: 230-234 (1985); Thomas &  
Capecchi, Cell 51: 503-512 (1987); Thompson et al., Cell 5:  
313-321 (1989); each of which is incorporated by reference  
10 herein in its entirety). For example, a mutant, non-  
functional LSG polynucleotide of the invention (or a  
completely unrelated DNA sequence) flanked by DNA  
homologous to the endogenous LSG polynucleotide sequence  
(either the coding regions or regulatory regions of the  
15 gene) can be used, with or without a selectable marker  
and/or a negative selectable marker, to transfect cells  
that express polypeptides of the invention *in vivo*. In  
another embodiment, techniques known in the art are used to  
generate knockouts in cells that contain, but do not  
20 express the gene of interest. Insertion of the DNA  
construct, via targeted homologous recombination, results  
in inactivation of the targeted gene. Such approaches are  
particularly suited in research and agricultural fields  
where modifications to embryonic stem cells can be used to  
25 generate animal offspring with an inactive targeted gene  
(e. g., see Thomas & Capecchi 1987 and Thompson 1989,  
*supra*). This approach can also be routinely adapted for  
use in humans provided the recombinant DNA constructs are  
directly administered or targeted to the required site *in*  
30 *vivo* using appropriate viral vectors that will be apparent  
to those of skill in the art.

In further embodiments of the invention, cells that  
are genetically engineered to express the LSG polypeptides  
of the invention, or alternatively, that are genetically  
35 engineered not to express the LSG polypeptides of the



- 91 -

invention (e. g., knockouts) are administered to a patient *in vivo*. Such cells may be obtained from the patient or a MHC compatible donor and can include, but are not limited to, fibroblasts, bone marrow cells, blood cells (e. g., lymphocytes), adipocytes, muscle cells, and endothelial cells. The cells are genetically engineered *in vitro* using recombinant DNA techniques to introduce the coding sequence of polypeptides of the invention into the cells, or alternatively, to disrupt the coding sequence and/or endogenous regulatory sequence associated with the polypeptides of the invention, e. g., by transduction (using viral vectors, and preferably vectors that integrate the transgene into the cell genome) or transfection procedures, including, but not limited to, the use of plasmids, cosmids, YACs, naked DNA, electroporation, liposomes, etc.

The coding sequence of the LSG polypeptides of the invention can be placed under the control of a strong constitutive or inducible promoter or promoter/enhancer to achieve expression, and preferably secretion, of the LSG polypeptides of the invention. The engineered cells which express and preferably secrete the LSG polypeptides of the invention can be introduced into the patient systemically, e.g., in the circulation, or intraperitoneally.

Alternatively, the cells can be incorporated into a matrix and implanted in the body, e.g., genetically engineered fibroblasts can be implanted as part of a skin graft or genetically engineered endothelial cells can be implanted as part of a lymphatic or vascular graft (see, for example, U.S. Patent 5,399,349 and U.S. Patent 5,460,959 each of which is incorporated by reference herein in its entirety).

When the cells to be administered are non-autologous or non-MHC compatible cells, they can be administered using well known techniques which prevent the development of a

- 92 -

host immune response against the introduced cells. For example, the cells may be introduced in an encapsulated form which, while allowing for an exchange of components with the immediate extracellular environment, does not  
5 allow the introduced cells to be recognized by the host immune system.

Transgenic and "knock-out" animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function  
10 of LSG polypeptides of the present invention, studying conditions and/or disorders associated with aberrant LSG expression, and in screening for compounds effective in ameliorating such LSG associated conditions and/or disorders.

15 The following nonlimiting example is provided to further illustrate the present invention.

#### EXAMPLE

The following Example is carried out using standard techniques, which are well known and routine to those of  
20 skill in the art, except where otherwise described in detail. Routine molecular biology techniques of the following example can be carried out as described in standard laboratory manuals, such as Sambrook et al., MOLECULAR CLONING: A LABORATORY MANUAL, 2nd Ed.; Cold  
25 Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989).

#### Introduction and background for Microarray analysis

cDNA microarrays are prepared by high-speed robotic printing of thousands of distinct cDNAs in an ordered array  
30 on glass microscope slides. They are used to measure the relative abundance of specific sequences in two complex samples (Schena et al, 1995; Shalon et al, 1996).

- 93 -

In the microarray procedure, mRNA is isolated from tissues of interest, either from a tumor or control (normal or normal adjacent tissue). mRNA (200-600 ng) from cancer tissue or control is reverse transcribed to incorporate the  
5 fluorescent nucleotides Cy5 (red) or Cy3 (green), respectively. The two populations of fluorescently labeled cDNA are mixed together and hybridized simultaneously to a microarray bearing approximately 10,000 cDNA elements in a 2cm x 2cm area on a glass slide (Microarrays hybridization  
10 service: Incyte Genomics, Fremont, CA, USA). After hybridization, the slides are scanned with a scanning laser confocal microscope.

The scanned image is used to generate the intensity and local background measurements for each spot on the  
15 array (GEMtools software, Incyte Genomics). For each spot, representing one EST, the ratio of the normalized Cy5/Cy3 intensities generates a quantitation of the gene's expression in one tissue relative to the control, in this case, the expression in cancer tissue versus either normal  
20 or normal adjacent tissue. For example, a gene that shows a Cancer-Cy5 intensity of 3000 and a Normal-Cy3 intensity of 1000 is expressed 3-fold more in cancer tissue. Advanced analysis software is used to sort and decipher patterns of gene expression from the data (Cluster and  
25 Treeview programs, Stanford University; Eisen et al, 1998; Alizadeh et al, 2000). However, the reproducibility study from Incyte shows that the level of detectable differential expression is calculated to be approximately plus or minus 1.74. Consequently, any elements with observed ratios  
30 greater than or equal to 1.8 between cancer and normal are deemed differentially expressed.

#### References:

1. Schena, M., D. Shalon, R.W. Davis, and P.O. Brown. 1995. Quantitative monitoring of gene expression patterns

- 94 -

with a complementary cDNA microarray. Science 270:  
467-470.

2. Shalon, D., S.J. Smith, and P.O. Brown. 1996. A DNA Microarray System for Analyzing Complex DNA samples Using Two-color Fluorescent Probe Hybridization. Genome Research 6: 639-645.
3. Eisen, M.B., P.T. Spellman, P.O. Brown, and D. Botstein. 1998. "Cluster analysis and display of genome-wide expression patterns". PNAS 95: 14863-14868.
- 10 4. Alizadeh, A.A., et al, 2000. "Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling." Nature, 403: 503-511.
5. GEM Microarray Reproducibility Study. Technical specifications from Incyte Genomics.

#### 15 Lung diaDexus microarray candidates

Following is a list of "diaDexus microarray candidates" sequences for lung cancer, also referred to herein as lung specific genes or LSGs:

	Sequences	Gene ID/ Clone ID/ ddxid
20	1	1040286/ 2746236/ 18867
	2	198406/ 2639142/ 12801
	3	441298/ 1877647/ 8255
	4	244318/ 3032060/ 7048
	5	429368/ 2890670/ 4002
25	6	975386/ 289582/ 5018
	7	480710/ 1911471/ 12153
	8	1040699/ 1899557/ 13678
	9	1040383/ 1556335/ 3273
	10	108494/ 3130429/ 3126
30	11	331878/ 2445607/ 3070
	12	233442/ 1959959/ 18837
	13	255993/ 1670828/ 7873
	14	897843 / 1823610/ 16315
	15	414885/ 2655867/ 21009

- 95 -

	16	1100375/ 690306/ x
	17	6133/ 3993331/ x
	18	257782/ 3032060A/ 7048A
	19	347005/ 1911471A/ 12153A
5	20	332710/ 3130429A/ 3126A
	21	255828/ 2445607A/ 3070A
	22	328565/ 3993331A/ x

Table 1 depicts numbers which are ratios indicating the levels of expression of the Clone IDs in the cancer tissue sample (labeled with Cy5) relative to the normal tissue, or the normal adjacent tissue control (labeled with Cy3) used in that experiment. The Cy5/Cy3 ratio of the normalized fluorescent intensities in each channel is used as a measure of relative gene expression. A positive number represents overexpression in cancer relative to the normal control. A negative number represents higher expression in the normal adjacent sample compared to the cancer tissue sample used in that experiment. X means no experiment was performed for the particular tissue sample.

20 Table 1:

	CloneID	LN.A143	LN.A160	LN.A182
		Vs. Apool	Vs. Apool	Vs. Apool
	2746236	5.3	2.6	1.5
	2639142	1.9	1.7	x
25	1877647	1.6	1.5	x
	3032060	2.1	1.2	x
	2890670	2.6	1.0	2.7
	289582	1.5	1.1	x
	CloneID	LN.A213	LN.A288	LN.A323
		Vs. Apool	Vs. Apool	Vs. Apool
30	2746236	4.6	8.9	3.7
	2639142	2.9	x	4.5
	1877647	2	x	5.0
	3032060	2.3	x	4.8
35	2890670	x	4.8	4.6
	289582	2.4	x	1.8
	CloneID	LN.A339	LN.A345	

- 96 -

	Vs. Apool	Vs. Apool
2746236	2.9	2.9
2639142	1.7	1.2
1877647	1.9	1.2
5 3032060	1.2	1.2
2890670	2.4	1.9
289582	2.4	1.9

Absolute values greater than or equal to 1.8 are considered to be above background levels, and are, therefore significant (Source: Incyte Genomics: GEM microarray technical specifications). The relative levels of expression in Table 1 show that Clone ID 2746236 mRNA expression is higher than background in 7 of the cancer tissue samples out of a total of 8 experiments. Clone ID 2639142 mRNA expression is higher than background in 3 of the cancer tissue samples out of a total of 6 experiments. Clone ID 1877647 mRNA expression is higher than background in 3 of the cancer tissue samples out of a total of 6 experiments. Clone ID 3032060 mRNA expression is higher than background in 3 of the cancer tissue samples out of a total of 6 experiments. Clone ID 2890670 mRNA expression is higher than background in 5 of the cancer tissue samples out of a total of 6 experiments. Clone ID 289582 mRNA expression is higher than background in 4 of the cancer tissue samples out of a total of 6 experiments.

An additional 16 clones have also been identified by the same type of experiments. These additional clones all show from 30% to 80 % overexpression in cancer tissue samples. The sequences of these LSGs are also disclosed herein.

### Semi-quantitative Polymerase Chain Reaction

Semi-quantitative Polymerase Chain Reaction (SQ-PCR) is a method that utilizes end point PCR on serial dilutions of cDNA samples in order to determine relative expression patterns of genes of interest in multiple samples. Using

- 97 -

random hexamer primed Reverse Transcription (RT) cDNA panels are created from total RNA samples. Gene specific primers are then used to amplify fragments using Polymerase Chain Reaction (PCR) technology from four 10x serial cDNA dilutions in duplicate. Relative expression levels of 0, 1, 10, 100 and 1000 are used to evaluate gene expression. A positive reaction in the most dilute sample indicates the highest relative expression value. This is determined by analysis of the sample reactions on a 2-4% agarose gel.

10 The tissue samples used include 12 normal, 12 cancer and 6 pairs tissue specific cancer and matching samples.

Of the list of "diaDexus microarray candidates" sequences for lung cancer, the following sequences were analyzed by semi-quantitative PCR and found to be

15 upregulated in lung adenocarcinoma/carcinoma.

Example#	SEQ ID NO:	Gene ID	Clone ID	ddxid	Sqlng code
	1	1040286	2746236	18867	Sqlng042
	2	441298	1877647	8255	Sqlng040
20	3	331878	2445607	3070	Sqlng046
	4	328565	3993331A	x	Sqlng050

Example 1 - SEQ ID NO:1

Semi quantitative PCR was done using the following primers:

25 Sqlng042 forward:

5' CCAGAGCCCAAATCTTGTGAC 3' (SEQ ID NO:23)

Sqlng042 reverse:

5' GCGGCTTTGTCTTGGCATT 3' (SEQ ID NO:24)

Table 2 shows absolute numbers which are relative levels of expression of Sqlng042 in 12 normal samples from 12 different tissues. These RNA samples are individual samples or are commercially available pools, originated by pooling samples of a particular tissue from different

30

- 98 -

individuals. Using Polymerase Chain Reaction (PCR) technology expression levels were analyzed from four 10x serial cDNA dilutions in duplicate. Relative expression levels of 0, 1, 10, 100 and 1000 are used to evaluate gene expression. A positive reaction in the most dilute sample indicates the highest relative expression value.

Table 2:

	Tissue	Normal
10	Breast	1000
	Colon	1000
	Endometrium	1000
	Kidney	1000
	Liver	10
15	Lung	1000
	Ovary	1000
	Prostate	100
	Small Intestine	1000
	Stomach	1000
20	Testis	1000
	Uterus	100

Relative levels of expression in Table 2 show that normal breast, colon, endometrium, kidney, lung, ovary, small intestine, stomach and testis show high expression of Sgln042. Moderate levels of expression are apparent in prostate and uterus. Low levels of expression are apparent in normal liver.

Table 3 shows absolute numbers which are relative levels of expression of Sgln042 in 12 cancer samples from 12 different tissues. Using Polymerase Chain Reaction (PCR) technology expression levels were analyzed from four 10x serial cDNA dilutions in duplicate. Relative expression levels of 0, 1, 10, 100 and 1000 are used to evaluate gene expression. A positive reaction in the most dilute sample indicates the highest relative expression value.

Table 3:

Tissue	Cancer
bladder	1000
breast	1000



- 99 -

	colon	1000
	kidney	1
	liver	100
	lung	1000
5	ovary	1
	pancreas	1000
	prostate	10
	stomach	1000
	testes	1
10	uterus	1000

Relative levels of expression in Table 3 show that Sqlng042 is expressed in low levels in kidney, ovary, and testis carcinomas. Sqlng042 is expressed in high levels in other tissue carcinomas.

15 Table 4 shows absolute numbers which are relative levels of expression of Sqlng042 in 6 lung cancer matching samples. A matching pair is formed by mRNA from the cancer sample for a particular tissue and mRNA from the normal adjacent sample for that same tissue from the same  
20 individual.

Using Polymerase Chain Reaction (PCR) technology expression levels were analyzed from four 10x serial cDNA dilutions in duplicate. Relative expression levels of 0, 1, 10, 100 and 1000 are used to evaluate gene expression.  
25 A positive reaction in the most dilute sample indicates the highest relative expression value

Table 4:

	Sample ID	Tissue	Cancer	NAT
	9702C115RB	lung	1	1
30	9502C032	lung	1000	1000
	8894A	lung	1	1000
	9704C060RA	lung	1	1
	11145B	lung	1	1000
	9502C109R	lung	1000	1000

35 Relative levels of expression in Table 4 show that Sqlng042 is expressed in high levels in two of the six lung cancer samples. However, high levels of expression was observed in the matching normal adjacent tissue (NAT).

- 100 -

**Example 2 - SEQ ID NO:3**

Semi quantitative PCR was done using the following primers:

Sq1ng040 forward:

5' 5' ATTGCCATCCCAGTGACAGTG 3' (SEQ ID NO:25)

Sq1ng040 reverse:

5' TTGGGAGATGTGGGTGATGAG 3' (SEQ ID NO:26)

Table 5 shows absolute numbers which are relative levels of expression of Sq1ng040 in 12 normal samples from 12 different tissues. These RNA samples are individual samples or are commercially available pools, originated by pooling samples of a particular tissue from different individuals. Using Polymerase Chain Reaction (PCR) technology expression levels were analyzed from four 10x serial cDNA dilutions in duplicate. Relative expression levels of 0, 1, 10, 100 and 1000 are used to evaluate gene expression. A positive reaction in the most dilute sample indicates the highest relative expression value.

**Table 5:**

	Tissue	Normal
20	Breast	0
	Colon	0
	Endometrium	1
	Kidney	0
25	Liver	0
	Lung	10
	Ovary	1
	Prostate	10
	Small Intestine	1
30	Stomach	1
	Testis	100
	Uterus	1

Relative levels of expression in Table 5 show that normal lung and prostate show moderate expression of Sq1ng040. High level expression is only apparent in testis. Low levels of expression are apparent in endometrium, ovary, small intestine and uterus.

- 101 -

Table 6 shows absolute numbers which are relative levels of expression of Sqlng040 in 12 cancer samples from 12 different tissues. Using Polymerase Chain Reaction (PCR) technology expression levels were analyzed from four 5 10x serial cDNA dilutions in duplicate. Relative expression levels of 0, 1, 10, 100 and 1000 are used to evaluate gene expression. A positive reaction in the most dilute sample indicates the highest relative expression value.

10 Table 6:

	Tissue	Cancer
	bladder	0
	breast	10
	colon	0
15	kidney	10
	liver	0
	lung	100
	ovary	100
	pancreas	100
20	prostate	10
	stomach	10
	testes	10
	uterus	10

Relative levels of expression in Table 6 show that Sqlng040 25 is expressed in moderate to high levels in breast, kidney, lung, ovary, pancreas, prostate, stomach, testis and uterus carcinomas.

Table 7 shows absolute numbers which are relative levels of expression of Sqlng040 in 6 lung cancer matching 30 samples. A matching pair is formed by mRNA from the cancer sample for a particular tissue and mRNA from the normal adjacent sample for that same tissue from the same individual. Using Polymerase Chain Reaction (PCR) technology expression levels were analyzed from four 10x 35 serial cDNA dilutions in duplicate. Relative expression levels of 0, 1, 10, 100 and 1000 are used to evaluate gene

- 102 -

expression. A positive reaction in the most dilute sample indicates the highest relative expression value.

Table 7:

	Sample ID	Tissue	Cancer	NAT
5	9702C115RB	lung	100	10
	9502C032	lung	100	1
	8894A	lung	10	0
	9704C060RA	lung	10	10
	11145B	lung	10	100
10	9502C109R	lung	100	10

Relative levels of expression in Table 7 show that Ssqlng040 is expressed in moderate levels in four of the six lung cancer samples compared with the expression in the matching normal adjacent tissue (NAT).

## 15 Example 3 - SEQ ID NO:11

Semi quantitative PCR was done using the following primers:

Ssqlng046 forward:

5' CCTGCCCTGGTATGTTTTTCTT 3' (SEQ ID NO:27)

## 20 Ssqlng046 reverse:

5' CAGCCACAAATGCCTTCTAC 3' (SEQ ID NO:28)

Table 8 shows absolute numbers which are relative levels of expression of Ssqlng046 in 12 normal samples from 12 different tissues. These RNA samples are individual samples or are commercially available pools, originated by pooling samples of a particular tissue from different individuals. Using Polymerase Chain Reaction (PCR) technology expression levels were analyzed from four 10x serial cDNA dilutions in duplicate. Relative expression levels of 0, 1, 10, 100 and 1000 are used to evaluate gene expression. A positive reaction in the most dilute sample indicates the highest relative expression value.

Table 8:

	Tissue	Normal
35	Breast	0

- 103 -

	Colon	10
	Endometrium	1
	Kidney	10
	Liver	1
5	Lung	10
	Ovary	10
	Prostate	0
	Small Intestine	0
	Stomach	0
10	Testis	10
	Uterus	1

Relative levels of expression in Table 8 show that normal colon, kidney, lung, and ovary show moderate expression of Sqlng046. Low levels of expression are apparent in endometrium and liver. No expression is apparent in other tissues.

Table 9 shows absolute numbers which are relative levels of expression of Sqlng046 in 12 cancer samples from 12 different tissues. Using Polymerase Chain Reaction (PCR) technology expression levels were analyzed from four 10x serial cDNA dilutions in duplicate. Relative expression levels of 0, 1, 10, 100 and 1000 are used to evaluate gene expression. A positive reaction in the most dilute sample indicates the highest relative expression value.

Table 9:

	Tissue	Cancer
	bladder	1
	breast	1
30	colon	0
	kidney	1
	liver	1
	lung	0
	ovary	0
35	pancreas	10
	prostate	0
	stomach	1
	testes	1
	uterus	1

Relative levels of expression in Table 9 show that Sqlng046 is expressed in low levels in bladder, breast, kidney,

- 104 -

liver, stomach, testis and uterus carcinomas. Sqlng046 is expressed in moderate levels only in pancreatic carcinoma.

Table 10 shows absolute numbers which are relative levels of expression of Sqlng046 in 6 lung cancer matching  
5 samples. A matching pair is formed by mRNA from the cancer sample for a particular tissue and mRNA from the normal adjacent sample for that same tissue from the same individual. Using Polymerase Chain Reaction (PCR) technology expression levels were analyzed from four 10x  
10 serial cDNA dilutions in duplicate. Relative expression levels of 0, 1, 10, 100 and 1000 are used to evaluate gene expression. A positive reaction in the most dilute sample indicates the highest relative expression value.

Table 10:

15	Sample ID	Tissue	Cancer	NAT
	9702C115RB	lung	10	10
	9502C032	lung	100	100
	8894A	lung	10	1
	9704C060RA	lung	10	10
20	11145B	lung	1	10
	9502C109R	lung	100	1

Relative levels of expression in Table 10 show that Sqlng046 is expressed in higher levels in two of the six lung cancer samples compared with the expression in  
25 matching normal adjacent tissue (NAT).

#### Example 4 - SEQ ID NO:22

Semi quantitative PCR was done using the following primers:

Sqlng050 forward:

30 5' CCACTAGGATTATTTCCAGCATAA 3' (SEQ ID NO:29)

Sqclng050 reverse:

5' GGTGTGAAAATATCTGGTCCACTT 3' (SEQ ID NO:30)

Table 12 shows absolute numbers which are relative levels of expression of Sqlng050 in 12 normal samples from

- 105 -

12 different tissues. These RNA samples are individual samples or are commercially available pools, originated by pooling samples of a particular tissue from different individuals. Using Polymerase Chain Reaction (PCR)

5 technology expression levels were analyzed from four 10x serial cDNA dilutions in duplicate. Relative expression levels of 0, 1, 10, 100 and 1000 are used to evaluate gene expression. A positive reaction in the most dilute sample indicates the highest relative expression value.

10 Table 12:

	Tissue	Normal
	Breast	100
	Colon	1000
	Endometrium	100
15	Kidney	100
	Liver	100
	Lung	100
	Ovary	1000
	Prostate	1000
20	Small Intestine	100
	Stomach	100
	Testis	10
	Uterus	100

Relative levels of expression in Table 12 show that normal  
 25 colon, ovary, and prostate show high expression of Ssqlng050. Moderate levels of expression are apparent in breast, endometrium, kidney, liver, lung, small intestine, stomach and uterus. Low levels of expression are apparent in normal testis

30 Table 13 shows absolute numbers which are relative levels of expression of Ssqlng050 in 12 cancer samples from 12 different tissues. Using Polymerase Chain Reaction (PCR) technology expression levels were analyzed from four 10x serial cDNA dilutions in duplicate. Relative  
 35 expression levels of 0, 1, 10, 100 and 1000 are used to evaluate gene expression. A positive reaction in the most dilute sample indicates the highest relative expression value.

- 106 -

Table 13:

	Tissue	Cancer
5	bladder	10
	breast	100
	colon	100
	kidney	10
	liver	100
10	lung	100
	ovary	100
	pancreas	100
	prostate	100
	stomach	100
	testes	100
	uterus	1000

15 Relative levels of expression in Table 13 show that  
 Sqlng050 is expressed in low to moderate levels in 11 out  
 of 12 different tissue carcinomas. Sqlng050 is only  
 expressed in high level in uterus carcinoma.

Table 14 shows absolute numbers which are relative  
 20 levels of expression of Sqlng050 in 6 lung cancer matching  
 samples. A matching pair is formed by mRNA from the cancer  
 sample for a particular tissue and mRNA from the normal  
 adjacent sample for that same tissue from the same  
 individual. Using Polymerase Chain Reaction (PCR)  
 25 technology expression levels were analyzed from four 10x  
 serial cDNA dilutions in duplicate. Relative expression  
 levels of 0, 1, 10, 100 and 1000 are used to evaluate gene  
 expression. A positive reaction in the most dilute sample  
 indicates the highest relative expression value.

30 Table 14:

	Sample ID	Tissue	Cancer	NAT
35	9702C115RB	lung	100	100
	9502C032	lung	1000	1000
	8894A	lung	100	1
	9704C060RA	lung	100	10
	11145B	lung	100	1000
	9502C109R	lung	100	10

Relative levels of expression in Table 14 show that  
 Sqlng050 is expressed in higher levels in three of the six



- 107 -

lung cancer samples compared to the expression level in the matching normal adjacent tissue (NAT).

#### Relative Quantitation of Gene Expression

Real-Time quantitative PCR with fluorescent Taqman probes is a quantitation detection system utilizing the 5'-3' nuclease activity of Taq DNA polymerase. The method uses an internal fluorescent oligonucleotide probe (Taqman) labeled with a 5' reporter dye and a downstream, 3' quencher dye. During PCR, the 5'-3' nuclease activity of Taq DNA polymerase releases the reporter, whose fluorescence can then be detected by the laser detector of the Model 7700 Sequence Detection System (PE Applied Biosystems, Foster City, CA, USA).

Amplification of an endogenous control is used to standardize the amount of sample RNA added to the reaction and normalize for Reverse Transcriptase (RT) efficiency. Either cyclophilin, glyceraldehyde-3-phosphate dehydrogenase (GAPDH) or 18S ribosomal RNA (rRNA) is used as this endogenous control. To calculate relative quantitation between all the samples studied, the target RNA levels for one sample were used as the basis for comparative results (calibrator). Quantitation relative to the "calibrator" can be obtained using the standard curve method or the comparative method (User Bulletin #2: ABI PRISM 7700 Sequence Detection System).

The tissue distribution and the level of the target gene were examined for every example in normal and cancer tissue. Total RNA was extracted from normal tissues, cancer tissues, and from cancers and the corresponding matched adjacent tissues. Subsequently, first strand cDNA was prepared with reverse transcriptase and the polymerase chain reaction was done using primers and Taqman probe specific to each target gene. The results are analyzed using the ABI PRISM 7700 Sequence Detector. The absolute

- 108 -

numbers are relative levels of expression of the target gene in a particular tissue compared to the calibrator tissue.

Example 1 - SEQ ID NO: 3

5 Table 15 shows absolute numbers which are relative levels of expression of the LSG of SEQ ID NO:3 in 24 normal different tissues. All the values are compared to normal small intestine (calibrator). These RNA samples are commercially available pools, originated by pooling samples  
10 of a particular tissue from different individuals.

Table 15:

	Tissue	NORMAL
	Adrenal Gland	0.56
	Bladder	0.03
15	Brain	2.57
	Cervix	0.42
	Colon	0.33
	Endometrium	5.12
	Esophagus	0.06
20	Heart	0.08
	Kidney	1.2
	Liver	1.38
	Lung	5.54
	Mammary Gland	3.96
25	Muscle	0.44
	Ovary	1.29
	Pancreas	7.94
	Prostate	5.21
	Rectum	1.36
30	Small Intestine	1
	Spleen	36.89
	Stomach	2.8
	Testis	10.16
	Thymus	179.15
35	Trachea	3.08
	Uterus	1.04
	0=negative	

The relative levels of expression in Table 15 show that mRNA expression of the LSG of SEQ ID NO:3 is very high  
40 in thymus (179.15) compared with all the other normal tissues analyzed. The expression level of the LSG of SEQ ID NO:3 is moderate in normal lung. Small intestine, the

- 109 -

calibrator, has a relative expression level of 1. These results demonstrated that mRNA expression of the LSG of SEQ ID NO:3 is relatively specific for lung.

The absolute numbers in Table 15 were obtained analyzing pools of samples of a particular tissue from different individuals. They can not be compared to the absolute numbers originated from RNA obtained from tissue samples of a single individual in Table 16.

Table 16 shows absolute numbers which are relative levels of expression of the LSG of SEQ ID NO:3 in 79 pairs of matching samples and 2 normal blood samples. All the values are compared to normal small intestine (calibrator). A matching pair is formed by mRNA from the cancer sample for a particular tissue and mRNA from the normal adjacent sample for that same tissue from the same individual.

Table 16:

	Sample ID	Cancer Type	Tissue	NORMAL	CANCER	MATCHING NORMAL ADJACENT
20	Lng60L	Adenocarcinoma	Lung 1		1.32	0.95
	Lng143L	Adenocarcinoma	Lung 2		9.29	0.96
	Lng60XL	Adenocarcinoma	Lung 3		41.5	13.18
	LngAC82	Adenocarcinoma	Lung 4		60.97	2.04
	LngAC88	Adenocarcinoma	Lung 5		50.21	31.89
25	LngAC66	Adenocarcinoma	Lung 6		1.42	0.72
	LngAC69	Adenocarcinoma	Lung 7		2.3	0.73
	LngAC11	Adenocarcinoma	Lung 8		2.41	1.95
	LngAC32	Adenocarcinoma	Lung 9		3.9	0.69
	LngAC94	Adenocarcinoma	Lung 10		2.65	0.77
30	LngAC90	Adenocarcinoma	Lung 11		16.85	0.57
	Lng223L	Adenocarcinoma	Lung 12		1.48	0.06
	LngAC39	Adenocarcinoma	Lung 13		139.1	1.52
	LngBR26	Bronchio-alveolar carcinoma	Lung 14		41.79	8.57
35	LngBA641	Bronchio-alveolar carcinoma	Lung 15		37.14	16
	LngSQ45	Squamous cell carcinoma	Lung 16		4.92	4.01
	LngSQ14	Squamous cell carcinoma	Lung 17		7.06	15.19
40	LngSQ9X	Squamous cell carcinoma	Lung 18		38.32	1.78
	LngSQ56	Squamous cell carcinoma	Lung 19		55.72	33.01
45	LngSQ80	Squamous cell carcinoma	Lung 20		34.42	4.3
	LngSQ32	Squamous cell carcinoma	Lung 21		69.55	21.86
50	LngSQ16	Squamous cell carcinoma	Lung 22		1.7	0.22

- 110 -

LngSQ79	Squamous cell carcinoma	Lung 23	4.71	3.04
Lng47XQ	Squamous cell carcinoma	Lung 24	35.26	1.42
5 LngBR94	Squamous cell carcinoma	Lung 25	138.62	0.19
LngC20X	Squamous cell carcinoma	Lung 26	3.05	0.18
10 LngSQ44	Squamous cell carcinoma	Lung 27	7.06	3.97
Lng90X	Squamous cell carcinoma	Lung 28	1.49	0.66
LngSQ43	Squamous cell carcinoma	Lung 29	97.01	1.71
15 LngLC71	Large cell carcinoma	Lung 30	27.86	16.22
LngLC109	Large cell carcinoma	Lung 31	102.89	20.25
LngLC80	Large cell carcinoma	Lung 32	34.66	10.13
20 Lng77L	Large cell carcinoma	Lung 33	1.03	9.22
Lng315L	Lung carcinoma	Lung 34	36.25	50.39
Lng528L	Lung carcinoma	Lung 35	21.48	6.54
25 Lng75XC	Metastatic from Osteogenic Sarcoma	Lung 36	3.53	4.55
LngMT67	Metastatic from renal cell cancer	Lung 37	8.2	3.97
LngMT71	Metastatic from melanoma	Lung 38	13.93	19.23
30 Bld46XK		Bladder 1	0	0
BldTR14		Bladder 2	1.57	0.78
B5		Blood 1	154.34	
B6		Blood 2	177.91	
35 CvxKS52		Cervix 1	11.96	2.27
CvxKS83		Cervix 2	92.09	8.66
ClnAS43		Colon 1	4.03	0.29
ClnAS45		Colon 2	0.28	0.17
ClnAS46		Colon 3	0.38	0.59
40 Cln AS67		Colon 4	0.62	1.78
Cln AS89		Colon 5	0.09	0.05
Endo28XA		Endometrium 1	15.51	4.77
Endo10479		Endometrium 2	24	7.14
Endo68X		Endometrium 3	13.13	14.42
45 Kid10XD		Kidney 1	3.07	2.07
Kid109XD		Kidney 2	8.22	7.24
Liv15XA		Liver 1	0.17	0.09
Liv174L		Liver 2	0.15	0.32
Mam355		Mammary 1	2.63	0.15
50 Mam173M		Mammary 2	6.87	7.67
Mam220		Mammary 3	0.29	0.87
Mam976M		Mammary 4	0.19	0.91
Ovr180B		Ovary 1	25.72	0
OvrA084		Ovary 2	2.7	1.97
55 Pan77X		Pancreas 1	8.11	3.25
Pan92X		Pancreas 2	27.28	21.78
Pro101XB		Prostate 1	6.99	4.68
Pro109XB		Prostate 2	1.42	1.16
Pro125XB		Prostate 3	2.24	1.71
60 Pro13XB		Prostate 4	0.41	1.59
Skn39A		Skin 1	3.71	0.35
Skn816S		Skin 2	25.81	0.34
SmInt21XA		Sm. Int. 1	4.35	1.17

- 111 -

	SmIntH89	Sm. Int. 2	13.93	3.16
	Sto115S	Stomach 1	4.59	5.17
	Sto264S	Stomach 2	6.39	4.16
	Sto288S	Stomach 3	5.01	0.46
	Thr270T	Thyroid 1	6.39	4.58
5	Thr939T	Thyroid 2	0.86	1.55
	Tst647T	Testis 1	2.49	0.43
	Tst663T	Testis 2	9.16	3.89
	Utr135XO	Uterus 1	0.34	0.43
10	Utr141XO	Uterus 2	2.51	0.63

In the analysis of matching samples, the higher levels of expression were in lung showing a high degree of tissue specificity for lung tissue. These results confirm the tissue specificity results obtained with normal pooled samples (Table 15).

Furthermore, the levels of mRNA expression in cancer samples and the isogenic normal adjacent tissue from the same individual were compared. This comparison provides an indication of specificity for the cancer stage (e.g. higher levels of mRNA expression in the cancer sample compared to the normal adjacent). Table 16 shows overexpression of the LSG of SEQ ID NO:3 in 26 lung cancer tissues compared with their respective normal adjacent (lung samples #2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 14, 15, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 29, 30, 31, and 32). There is overexpression in the cancer tissue for 68% of the lung matching samples tested (total of 38 lung matching samples).

Altogether, the relative high level of lung tissue specificity, plus the mRNA overexpression in 68% of the lung carcinoma matching samples tested are believed to make the LSG of SEQ ID NO:3 a good diagnostic marker for lung cancer.

Primers used for expression analysis are:

Forward

5' AGCCATTGCCATCCAGT 3' (SEQ ID NO:31)

Reverse

5' ATGTTCTTCACGCTCTTCGC 3' (SEQ ID NO:32)

Probe

- 112 -

5' AGGAAGTGCTGGAAGAGGCTGGCT 3' (SEQ ID NO:33)

### Example 2 - SEQ ID NO: 15

Table 17 shows absolute numbers which are relative levels of expression of the LSG of SEQ ID NO:15 in 24 normal different tissues. All the values are compared to normal brain (calibrator). These RNA samples are commercially available pools, originated by pooling samples of a particular tissue from different individuals.

Table 17:

10	<b>Tissue</b>	<b>NORMAL</b>
	Adrenal Gland	67.65
	Bladder	39.67
	Brain	1.00
	Cervix	677.93
15	Colon	1287.18
	Endometrium	162.58
	Esophagus	1034.70
	Heart	4.81
	Kidney	25.02
20	Liver	194.01
	Lung	4705.07
	Mammary Gland	840.44
	Muscle	12.91
	Ovary	608.87
25	Pancreas	20.89
	Prostate	858.10
	Rectum	4435.87
	Small Intestine	2149.82
	Spleen	5595.30
30	Stomach	14115.57
	Testis	64.67
	Thymus	2187.40
	Trachea	2866.35
	Uterus	193.34
35	0=negative	

The relative levels of expression in Table 17 show that mRNA expression of the LSG of SEQ ID NO:15 is very high in stomach (14115.57) compared with all the other normal tissues analyzed. Expression levels of this LSG are moderate in normal lung (4705.07) Brain, the calibrator, has a relative expression level of 1. These results

- 113 -

demonstrate that mRNA expression of the LSG of SEQ ID NO:15 is relatively specific for lung.

The absolute numbers in Table 17 were obtained analyzing pools of samples of a particular tissue from 5 different individuals. They can not be compared to the absolute numbers originated from RNA obtained from tissue samples of a single individual in Table 18.

Table 18:

	Sample ID	Cancer Type	Tissue	NORMAL	CANCER	MATCHING NORMAL ADJACENT
10	Lng60L	Adenocarcinoma	Lung 1		18561.17	5732.70
	Lng143L	Adenocarcinoma	Lung 2	28.54		1.57
	LngAC66	Adenocarcinoma	Lung 3	16555.24		3408.69
	LngAC69	Adenocarcinoma	Lung 4	18116.29		1891.09
15	LngAC11	Adenocarcinoma	Lung 5	4389.98		5732.70
	LngAC32	Adenocarcinoma	Lung 6	18179.19		10015.87
	LngAC94	Adenocarcinoma	Lung 7	10623.71		309.76
	Lng223L	Adenocarcinoma	Lung 8	8393.17		491.14
20	LngBR26	Bronchio-alveolar carcinoma	Lung 9	13.98		20.68
	LngBA641	Bronchio-alveolar carcinoma	Lung 10	34.78		10.13
	LngSQ45	Squamous cell carcinoma	Lung 11	9184.59		8995.58
25	LngSQ14	Squamous cell carcinoma	Lung 12	2.82		32.11
	LngSQ80	Squamous cell carcinoma	Lung 13	68.12		4.07
30	LngSQ16	Squamous cell carcinoma	Lung 14	3373.43		86.22
	LngSQ79	Squamous cell carcinoma	Lung 15	19215.37		81245.48
	Lng90X	Squamous cell carcinoma	Lung 16	5.19		1.14
35	LngSQ43	Squamous cell carcinoma	Lung 17	24.17		2.12
	LngLC71	Large cell carcinoma	Lung 18	67.42		25.37
40	LngLC109	Large cell carcinoma	Lung 19	12.38		3.96
	LngMT71	Metastatic from melanoma	Lung 20	13.00		9.45
	Bld46XK		Bladder 1	131.60		5.90
45	BldTR14		Bladder 2	8306.36		7009.03
	CvxKS52		Cervix 1	24.85		8.91
	ClnAS43		Colon 1	1590.21		8335.19
	ClnAS45		Colon 2	1458.23		1820.35
	ClnAS46		Colon 3	2418.67		3019.30
50	ClnAS67		Colon 4	365.82		823.14
	ClnAS89		Colon 5	2304.12		75.32
	Endo28XA		Endometrium 1	10.70		0.49
	Kid10XD		Kidney 1	0.38		0.21
	Liv15XA		Liver 1	19.16		115.76
55	Mam355		Mammary 1	16.56		0.18
	Pan77X		Pancreas 1	0.15		0.07

- 114 -

Pro101XB	Prostate 1		
Skn816S	Skin 1	2.46	1.05
SmInt21XA	Sm. Int. 1	0.28	0.10
Sto288S	Stomach 1	6.43	12.04
5 Thr270T	Thyroid 1	7.41	14.32
Tst647T	Testis 1	0.99	0.14
Utr135XO	Uterus 1	1217.75	15.62
0=negative		237.21	55.14

In the analysis of matching samples, the level of mRNA expression in cancer samples and the isogenic normal adjacent tissue from the same individual were compared. This comparison provides an indication of specificity for the cancer stage (e.g. higher levels of mRNA expression in the cancer sample compared to the normal adjacent). Table 18 shows overexpression of the LSG of SEQ ID NO:15 in 14 lung cancer tissues compared with their respective normal adjacent (lung samples #1, 2, 3, 4, 6, 7, 8, 10, 13, 14, 16, 17, 18, and 19). There is overexpression in the cancer tissue for 70% of the lung matching samples tested (total of 20 lung matching samples).

Altogether, the relative high level of lung tissue specificity, plus the mRNA overexpression in 70% of the lung carcinoma matching samples tested are believed to make the LSG of SEQ DI NO:15 a good diagnostic marker for lung cancer.

Primers used for expression analysis in this example are as follows:

Forward

5' AAGGGAGCACCGTGGAGAA 3' (SEQ ID NO:34)

30 Reverse

5' AGGGCTGGATGACTTGGGA 3' (SEQ ID NO:35)

Probe

5' TTCCCAACTCTAACCCACCCACG 3' (SEQ ID NO:36)



- 115 -

What is claimed is:

1. A LSG comprising:
  - (a) a polynucleotide of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 37 or 38 or a variant thereof;
  - (b) a polypeptide expressed by a polynucleotide of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 37 or 38 or a variant thereof; or
  - (c) a polynucleotide which is capable of hybridizing under stringent conditions to the antisense sequence of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 37 or 38.
2. The LSG of claim 1 wherein the polypeptide comprises SEQ ID NO: 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, or 56.
3. A method for diagnosing the presence of lung cancer in a patient comprising:
  - (a) determining levels of a LSG of claim 1 in cells, tissues or bodily fluids in a patient; and
  - (b) comparing the determined levels of LSG with levels of LSG in cells, tissues or bodily fluids from a normal human control, wherein a change in determined levels of LSG in said patient versus normal human control is associated with the presence of lung cancer.
4. A method of diagnosing metastases of lung cancer in a patient comprising:
  - (a) identifying a patient having lung cancer that is not known to have metastasized;
  - (b) determining levels of a LSG of claim 1 in a sample of cells, tissues, or bodily fluid from said patient; and

- 116 -

(c) comparing the determined LSG levels with levels of LSG in cells, tissue, or bodily fluid of a normal human control, wherein an increase in determined LSG levels in the patient versus the normal human control is associated  
5 with a cancer which has metastasized.

5. A method of staging lung cancer in a patient having lung cancer comprising:

- (a) identifying a patient having lung cancer;
- (b) determining levels of a LSG of claim 1 in a  
10 sample of cells, tissue, or bodily fluid from said patient;  
and

(c) comparing determined LSG levels with levels of LSG in cells, tissues, or bodily fluid of a normal human control, wherein an increase in determined LSG levels in  
15 said patient versus the normal human control is associated with a cancer which is progressing and a decrease in the determined LSG levels is associated with a cancer which is regressing or in remission.

6. A method of monitoring lung cancer in a patient  
20 for the onset of metastasis comprising:

- (a) identifying a patient having lung cancer that is not known to have metastasized;
- (b) periodically determining levels of a LSG of claim 1 in samples of cells, tissues, or bodily fluid from said  
25 patient; and

(c) comparing the periodically determined LSG levels with levels of LSG in cells, tissues, or bodily fluid of a normal human control, wherein an increase in any one of the periodically determined LSG levels in the patient versus  
30 the normal human control is associated with a cancer which has metastasized.

- 117 -

7. A method of monitoring a change in stage of lung cancer in a patient comprising:

- (a) identifying a patient having lung cancer;
- (b) periodically determining levels of a LSG of claim 1 in cells, tissues, or bodily fluid from said patient; and
- (c) comparing the periodically determined LSG levels with levels of LSG in cells, tissues, or bodily fluid of a normal human control, wherein an increase in any one of the periodically determined LSG levels in the patient versus the normal human control is associated with a cancer which is progressing in stage and a decrease is associated with a cancer which is regressing in stage or in remission.

8. A method of identifying potential therapeutic agents for use in imaging and treating lung cancer comprising screening compounds for an ability to bind to or decrease expression of a LSG of claim 1 relative to the LSG in the absence of the compound wherein the ability of the compound to bind to the LSG or decrease expression of the LSG is indicative of the compound being useful in imaging and treating lung cancer.

9. An antibody which specifically binds a polypeptide encoded by a LSG of claim 1.

10. The antibody of claim 9 wherein the polypeptide comprises SEQ ID NO:39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55 or 56.

11. A method of imaging lung cancer in a patient comprising administering to the patient an antibody of claim 9.

12. The method of claim 11 wherein said antibody is labeled with paramagnetic ions or a radioisotope.

- 118 -

13. A method of treating lung cancer in a patient comprising administering to the patient a compound which downregulates expression or activity of a LSG of claim 1.

14. A method of inducing an immune response against  
5 a target cell expressing a LSG of claim 1 comprising delivering to a human patient an immunogenically stimulatory amount of a LSG polypeptide so that an immune response is mounted against the target cell.

15. The method of claim 14 wherein the LSG  
10 polypeptide comprises SEQ ID NO:39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55 or 56.

16. A vaccine for treating lung cancer comprising a LSG of claim 1.

## SEQUENCE LISTING

<110> Macina, Roberto A  
 Nair, Manoj  
 Chen, Seiyu  
 diaDexus, Inc.

<120> Compositions and Methods Relating to Lung Specific  
 Genes

<130> DEX-0215

<140>

<141>

<150> 60/219,834

<151> 2000-07-21

<160> 56

<170> PatentIn Ver. 2.1

<210> 1

<211> 1449

<212> DNA

<213> Homo sapiens

<400> 1

```

ctgggagtggtg atttcataca ttctgtggagg tgggtgagagg atataactacg cagactctgt 60
gagggggccga ttcaccgtct ccagggacaa cgccaagaac tcactctatc tgcaaatgaa 120
cagcctgaga gccgaggaca cggctgttta tttctgtgag agagagccac cagcacccaa 180
ttactttgac tgctggagcc agggaaacct ggtcaccgtc tcctcagctt ccaccaaggg 240
cccatcggtc ttccccctgg cgccctgctc caggagcacc tctgggggca cagcggccct 300
gggctgcctg gtcaaggact acttccccga accggtgacg gtgtcgtgga actcaggcgc 360
cctgaccagc ggcgtgcaca ccttcccggc tgtcctacag tcctcaggac tctactccct 420
cagcagcgtg gtgaccgtgc cctccagcag cttgggcacc cagacctaca cctgcaacgt 480
gaatcacaag ccagcaaca ccaagggtga caagagagtt gagctcaaaa cccacttgg 540
tgacacaact cacacatgcc cacggtgccc agagcccaaa tcttgtgaca cacctcccc 600
atgccacagg tgcccagagc ccaaactctg tgacacacct ccccatgcc cacggtgccc 660
agcacctgaa ctctgaggag gaccgtcagt cttcctcttc ccccaaaac ccaaggatac 720
ccttatgatt tcccggaccc ctgaggtcac gtgcgtggtg gtggacgtga gccacgaaga 780
ccccgaggtc cagttcaagt ggtacgtgga cggcgtggag gtgcataatg ccaagacaaa 840
gccgcgggag gagcagttca acagcacgtt ccgtgtggtc agcgtcctca ccgtcctgca 900
ccaggactgg ctgaacggca aggagtacaa gtgcaagggtc tccaacaaag ccctcccagc 960
cccatcgag aaaaccatct ccaaaaccaa aggacagccc cgagaaccac aggtgtacac 1020
cctgccccca tcccgggagg agatgaccaa gaaccagggtc agcctgacct gcctgggtcaa 1080
aggcttctac ccagcgaca tcgccgtgga gtgggagagc agcgggcagc cggagaacaa 1140
ctacaacacc acgctccca tgctggactc cgacggctcc ttcttctct acagcaagct 1200

```



caccgtggac aagagcaggt ggcagcaggg gaacatcttc tcatgctccg tgatgcatga 1260  
 ggctctgcac aaccgcttca cgcagaagag cctctccctg tctccgggta aatgagtgcg 1320  
 acggccggca agcccccgct ccccgggctc tcggggtcgc gcgaggatgc ttggcacgta 1380  
 ccccggtgtac atacttcccc ggcacccagc atggaaataa agcacccagc gctgccctgg 1440  
 gccctgcg 1449

<210> 2

<211> 3825

<212> DNA

<213> Homo sapiens

<220>

<221> unsure

<222> (428)

<220>

<221> unsure

<222> (430)

<220>

<221> unsure

<222> (1997) .. (1998)

<400> 2

ttagtcacgt gaaaacccat caggaaatta aacttgatga tagcaacatt cctcctccct 60  
 ctttaaaaac acgcccaccg tcaccaactt ttatcacaat agaatctact gcccgcagaa 120  
 cagaaaaccc tactaagaac gagctttctc agtcccctaa aaaggacagt tatgttgaaac 180  
 ccccaccaag aaggcccatg tcgcaaaaat ctgaaattca cagagcaaac acttcccctt 240  
 ctccaccag gagtcgctct gaacaacttg tcagactcaa agacaccact gcaaagttat 300  
 ccaaaggggc catcccatgt ccagcagcaa ccccggttcc aattgtagag aagaggctctg 360  
 aaatcatcat gtctcctgca acacttcgtc gtcaaattaa gatagaaact cgtggtaggg 420  
 actctccnan ctacaatcac aataccagta aatataaatc atgctgctag tggttccttc 480  
 agagaatctg tggacgctca agaggaaatc aggaaagtgg agaagagagc tacttatgtt 540  
 cataaagatg gactaaattc cactgatcac atgggtgccc acactgaaag ttatgatgca 600  
 gttgaaatca tccgcaaggt tgcagtgcct cctgcctgt cagagcacac acagagatat 660  
 gaagcggcca accgaactgt tcaaatggct gaaaatttcg tgaatgacct tgaaaatgaa 720  
 ataaacagat gggttcaggga atttgagcat ggcccagttt ctgaagcaaa gtcaaataga 780  
 agagtttatg caaagggaga aacaaacccat aacatacaac aagaaagtcg tacatthttgt 840  
 aaggaggaat ttggattaac atcttttagga aacacgagtt ttacagactt ttcttgcaaa 900  
 catcctagag aactgcgaga aaagattcct gttaagcagc ccaggatctg ctctgaaacc 960  
 aggtctctaa gtgaacattt ctccaggcatg gatgcatttg agagtcaaat tggtagtgcg 1020  
 aagatgaaaa cctcttcac acatagctca gaagctggca aatctggctg tgacttcaag 1080  
 catgccccac caacctatga ggatgtcatt gctggacata ttttagatat ctctgattca 1140  
 cctaaagaag taagaaaaaa ttttcaaaag acgtggcaag agagtggaag agttttttaa 1200  
 ggcttgggat atgcaaccgc agatgcttct gcaactgaga tgagaaccac cttccaagag 1260  
 gaatctgcat ttataagtga agctgctgct ccaagacaag gaaatatgta tactttgtca 1320  
 aaagacagtt tatccaatgg agtgcctagt ggcagacaag cagaattttc ataagtcctg 1380





```

cttccgatgc caccattgca acagtaaact aagtttggga aattatgcat cacttcatgg 1440
acaaatatac tgtaaacctc acttttaaca acttttcaaa tccaaaggaa attatgatga 1500
aggttttgga cataagcagc ataaagatag atggaactgc aaaaacccaaa gcagatcagt 1560
ggacttttatt cctaatagaag aaccaaatat gtgtaaaaat attgcagaaa acacccttgt 1620
acctggagat cgtaatagaac atttagatgc tggtaacagt gaagggcaaa ggaatgattt 1680
gagaaaaatta ggggaaagg gaaaattaaa agtcatttgg cctccttcca aggagatccc 1740
taagaaaacc ttaccctttg aggaagagct caaaatgagt aaacctaagt ggccacctga 1800
aatgacaacc ctgctatccc ctgaatttaa aagtgaatct ctgctagaag atgttagaac 1860
tccagaaaat aaaggacaaa gacaagatca ctttccattt ttgcagcctt atctacagtc 1920
cacccatgtt tgtcagaaag aggatgttat aggaatcaaa gaaatgaaaa tgcctgaagg 1980
aagaaaagat gaaaagnngg aaggaaggaa gaatgtgcaa gataggccga gtgaagctga 2040
agacacaaaag agtaacagga aaagtgtctat ggatcttaat gacaacaata atgtgattgt 2100
gcagagtgtc gaaaaggaga aaaatgaaaa aactaaccaa actaatggtg cagaagtttt 2160
acaggttact aacactgatg atgagatgat gccagaaaat cataaagaaa atttgaataa 2220
gaataataat aacaattatg tagcagtcctc atatctgaat aattgcaggc agaagacatc 2280
tatttttagaa tttcttgatc tattaccctt gtcgagtga gcaaatagaca ctgcaaata 2340
atatgaaatt gagaagttag aaaatacatc tagaatctca gagttacttg gtatatttga 2400
atctgaaaag acttattcga ggaatgtact agcaatggct ctgaagaaac agactgacag 2460
agcagctgct ggcagtcctg tgcagcctgc tccaaaacca agcctcagcc agaggcctta 2520
tggtaaagg ggggaagtta atcatctctc ctgatacaaa tctcttaaac attaaaggaa 2580
gccattcaaa gagcaaaaat ttacactttt tcttttctaa caccgtgaaa atcactgcat 2640
tttccaagaa aaatgagaac attttcaatt gtgatttaat agattctgta gatcaaat 2700
aaaatatgcc atgcttggat ttaagggaat ttggaaagga tgtaaacct tggcatgttg 2760
aaacaacaga agctgcccgc aataatgaaa acacagggtt tgatgctctg agccatgaat 2820
gtacagctaa gcctttgttt cccagagtgg aggtgcagtc agaacaactc acggtggaag 2880
agcagattaa aagaaacagg tgctacagt acactgagta aaatatctat ggccactgac 2940
agtccacact taggcactga gagatattga tgttctgaaa taagatttta tgaatttga 3000
tacccttttg aggaacttga tgtaaacatg gtgttcagaa atctcgtgtc tatctcaatg 3060
ggatatttct tgtattacac cttgtcattt ttttcacaat ttatttacat ctacttttgt 3120
ttgaactgga atgaagagat gaaacactat ggatatgttt tccattcaaa tggcacttta 3180
gcatattgtt ctgttttctt gtaaaacatc atgggtgtga tttttatact gctgctgctt 3240
gtcacaatta ttataacttc tctgtaattt cctctgaaat aaaattgaat cacctgaggt 3300
gcaaaccaaa atacttctgt aacttttttt gatataact gtcattctaa gtacatatac 3360
tccttgtgac ttgggaagta tttgtcttga ggcaagtatt taccaccac actaaaataa 3420
tgctggaaaa aataaaatac taaactgaag gcacagtatt attagaaagt gtaacatttt 3480
cattttctct tttactccac attttaaaga tacgagggtt attgttcttg aaataattac 3540
ctatattaaa ttatcataga atgtgtctat aaacatttga cgaaaaatgt tgattttcct 3600
ccagaataat gtgaagtcca tactcagaaa ttaactagaa aggtttttaga cattacttaa 3660
ataaattatt cacattgcat ttgtattgct tgcctgtgtt aatggataag tataacaatc 3720
atatcactac agtttgtcag gttttcttct tatcatattt gatgaatatt aagtttttct 3780
gttatgaaaa catattcctc taaaatttgg cttctaaatt ttcta 3825

```

&lt;210&gt; 3

&lt;211&gt; 2315

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens



&lt;400&gt; 3

gtaagcagca gttgattaga attaaatgag cttgaatttg attctgacat tcatattgat 60  
 ttgtccttcc ctcaaaaaac accctgagta tggacagggc ttcccgactc tgcagactac 120  
 acgccgtcca tgagcagtgc ccagggtgtca ttacctgccc atgagatgtg acctgggcag 180  
 ggggtcccccac ctgtaccctt gggccccagg aggggaagccc agcatgtcag gctgaagcgg 240  
 ggggtgcttcc agagatggcc atgcagagca gccctcccgc ctcggttcct gaggccccgc 300  
 tcagtgggtcc cccactctg cagaatgtgc acccccagct ctgatgtctc ttccagggtga 360  
 aatccgggtc cccggccgtg ctggcattcg caaaggagaa gtcttttggg tggcccagct 420  
 tcatcacata cacggtcggc gtctcggacc ccggggtgg cagccaaggg cctctgtcca 480  
 ctacctgac cttctccagc cccgtgacca accaagccat tgccatccca gtgacagtgg 540  
 cttttgtgat ggatcgccgt gggcccggtc cttatggagc cagcctcttc cagcacttcc 600  
 tggattccta ccagggtcatg ttcttcacgc tcttcgccct gttgggtggg acagcgggtca 660  
 tgatcatagc ctaccacact gtctgcacgc cccgggatct tgctgtgcct gcagccctca 720  
 cgcctcgagc cagccctggg acacagcccc cactatttcg ctgectcacc acccacatct 780  
 cccaatgcat tgctcctgc tcgcaaagcc agccctccct cagggtgtg gagcccaggc 840  
 ctatggcctc ccactagggc cgcgtgaagg ttcccggagg atggggtctc agccgagcct 900  
 cgttgcaacc cccaagatgg aacatccctt gctgcattca cactggaaca agcccctcca 960  
 gatgagtgcc ccggccccag gccagcttca ctgcccgtctc ttcacacaga gctgtagtgt 1020  
 cggtcttgcc cattagctca ttttatgtag gagttttaaa tgtgtgtttt tttcctttca 1080  
 agtcttaca agctaagact ttttggtcctc ttcctttttg catggttgtc tagggtttct 1140  
 ggacaatgtg ctgttgcat tttattttcc tagccttgc aaaatcttcc ccttctcaag 1200  
 actttgagca gttagaagt ctcttttaga gttgtctgtg ggtgatgtta ctgtagtgg 1260  
 ctcagggaaa ggattgtcca gttactttag ggggtttttg gtgggggtttt tccccctgtg 1320  
 aaaacttact ttgcccctag tctgggtgct gctaggactt ctgaggagca atgggacatg 1380  
 agtgtccctg tatctgcgc actgcccga ggggaagcctc aggaaccagc acctggaggc 1440  
 caggatagcc aagccctggg tgagcgagag gctggagaac acaggagctc acccagggtc 1500  
 gctgcccac catgggccac tgtgaacaga cttcagtcct ctgtttttgt ttcataagcc 1560  
 gttgagacat ctgatggact tggetttaggc cctgctggga catcccacgt gtgatccctt 1620  
 tcactccac aggacaccag gactgtcctt aggaaaatgt ccttgagatg gcagcaggag 1680  
 tcatattttc tgtgtgtgtg tttcggaag ccgctgtgtc ctgcccagc acaaagaccc 1740  
 agtgtcattt gctcctcctg ttcctgtgac actccagaac ctgagcagat ctgagccacc 1800  
 gcctgccagt gtgagaggcg gccactttca tggcagctca tcaggcgagc gggcccagac 1860  
 agcttcccag caggccctag agcccggcct gggccaatga tggagggcgg ccgccagccc 1920  
 agggcctgcc catccagaag ggactcccca gggcctgggg gaggagacc ttggaaaagt 1980  
 cctctcttcc cagctcctga ttctggatct gagattctca gatcacaggc ccctgtgctc 2040  
 caggccgagg ctgggctacc ctgagggaga tccagagact catgcccatg gccatccatg 2100  
 cgtggacgct gtgtggagag tccaggatga cgggatcccg cacaagctcc cttcagtcct 2160  
 tcagggtgg gccatgtggg tgatttttct aaagctggag aaaggaagaa ttgtgccttg 2220  
 catattactt gagcttaaac tgacaacctg gatgtaaata ggagccttcc tactgggtta 2280  
 ttaataaag ttctatgtga tttttaaaaa aaaaaa 2315

&lt;210&gt; 4

&lt;211&gt; 300

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;



&lt;221&gt; unsure

&lt;222&gt; (8)

&lt;400&gt; 4

```

agcatganaa aggtgaaggc tgccgggtggc acgggggctcg gatctgctgc cgggccgacc 60
tgggagagcc atgaggctgt atgtgatggg gcacctcttg ggtgcacact ttggatgaca 120
agtgccccca agaggagcca gggctggctg cagtgaggcc ccaggagggt ctcagggggc 180
gtcctgcttc agctcaaggg gctaggaata ggggaaacga tgcagggaag ccaatggccc 240
aagtggctcc ctactgact gttacttget gtgtatgtct ctttcttttc ttttttttcc 300

```

&lt;210&gt; 5

&lt;211&gt; 4347

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 5

```

gcgggtgcggc ggccgggaggc ggaggcgagg gtgcgatggc gcggagcccg ggacgcgcgt 60
acgccctgct gcttctcctg atctgcttta acgttggaag tggacttcac ttacaggctct 120
taagcacaag aaatgaaaat aagctgcttc ctaaaccatcc tcatttagtg cggcaaaagc 180
gcgcctggat caccgcccc gtggctcttc gggaggggaga ggatctgtcc aagaagaatc 240
caattgccaa gatacattct gatcttgacg aagaaagagg actcaaaatt acttataaat 300
acactggaaa agggattaca gagccacctt ttgggtatatt tgtctttaac aaagatactg 360
gagaactgaa tgttaccagc attcttgatc gagaagaaac accatttttt ctgctaacag 420
gttacgcttt ggatgcaaga ggaaacaatg tagagaaacc cttagagcta cgcattaagg 480
ttcttgatat caatgacaac gaaccagtgt tcacacagga tgtctttggt ggggtctggtg 540
aagagttgag tgcagcacat actcttgatg tgaaaatcaa tgcaacagat gcagatgagc 600
ccaataccct gaattcgaaa atttcctata gaatcgatc tctggagcct gcttatcctc 660
cagtgttcta cctaaataaa gatacaggag agatttatac aaccagtgtt accttgga 720
gagaggaaca cagcagctac actttgacag tagaagcaag agatggcaat ggagaagtta 780
cagacaaacc tgtaaaacaa gctcaagttc agattcgat tttggatgtc aatgacaata 840
tacctgtagt agaaaataaa gtgcttgaag ggatgggtga agaaaatcaa gtcaacgtag 900
aagttacgcg cataaaagtg ttcgatgcag atgaaatagg ttctgataat tggctggcaa 960
attttacatt tgcacagga aatgaaggag gttatttcca catagaaaca gatgctcaaa 1020
ctaacgaagg aattgtgacc cttattaagg aagtagatta tgaagaaatg aagaatcttg 1080
acttcagtgt tattgtcgct aataaagcag cttttcaciaa gtcgattagg agtaaatata 1140
agcctacacc cattcccatc aaggctcaaag tgaaaaatgt gaaagaaggc attcatttta 1200
aaagcagcgt catctcaatt tatgttagcg agagcatgga tagatcaagc aaaggccaaa 1260
taattggaaa ttttcaagct tttgatgagg aactggact accagcccat gcaagatatg 1320
taaaattaga agatagagat aattggatct ctgtggattc tgtcacatct gaaattaaac 1380
ttgcaaaact tcctgatttt gaatctagat atgttcaaaa tggcacatac actgtaaaga 1440
ttgtggccat atcagaagat tctctagaa aaaccatcac tggcacagtc cttatcaatg 1500
ttgaagacat caacgacaac tgtccacac tgatagagcc tgtgcagaca atctgtcacg 1560
atgcagagta tgtgaatgtt actgcagagg acctggatgg acacccaaac agtggccctt 1620
tcagtttctc cgtcattgac aaaccacctg gcatggcaga aaaatggaaa atagcacgcc 1680
aagaaagtac cagtgtgctg ctgcaacaaa gtgagaaaaa gcttgggaga agtgaaattc 1740
agttcctgat ttcagacaat cagggtttta gttgtcctga aaagcaggtc cttacactca 1800
cagtttgtga gtgtctgcat ggcagcggct gcagggaagc acagcatgac tcctatgtgg 1860

```



```

gcctgggacc cgcagcaatt ggcgtcatga ttttggcctt tctgctcctg ctattggtac 1920
cacttttact gctgatgtgc cattgcggaa agggcgccaa aggcctttacc cccatacctg 1980
gcaccataga gatgctgcat ccttggaata atgaaggagc accacctgaa gacaaggtgg 2040
tgccatcatt tctgccagt gatacaagggg gcagctctagt aggaagaaat ggagtaggag 2100
gtatggccaa ggaagccacg atgaaaggaa gtagctctgc ttccattgtc aaagggcaac 2160
atgagatgtc cgagatggat ggaaggtggg aagaacacag aagcctgctt tctggtagag 2220
ctaccagtt tacaggggccc acaggcgcta tcatgaccac tgaaaccacg aagaccgcaa 2280
gggcccacag ggcttccaga gacatggccg gagctcaggc agctgctgtt gcactgaacg 2340
aagaattctt aagaaattat ttcactgata aagcggcctc ttacactgag gaagatgaaa 2400
atcacacagc caaagattgc cttctgggtt attctcagga agaaactgaa tctgtgaatg 2460
cttctattgg ttgttgacgt tttattgaag gagagctaga tgaccgcttc ttagatgatt 2520
tgggacttaa attcaagaca ctagctgaag tttgcctggg tcaaaaaata gatataaata 2580
aggaaattga gcagagacaa aaacctgcca cagaaacaag tatgaacaca gcttcacatt 2640
cactctgtga gcaactatg gttaattcag agaataccta ctctctggc agtagcttcc 2700
cagttccaaa atctttgcaa gaagccaatg cagagaaagt aactcaggaa atagtactg 2760
aaagatctgt gtcttctagg caggcgcaa aggtagctac acctcttctt gacccaatgg 2820
cttctagaaa tgtgatagca acagaaactt cctatgtcac aggttccact atgccaccaa 2880
ccactgtgat cctgggtcct agccagccac agagccttat tgtgacagag aggtgtgatg 2940
ctccagcttc taccttggtg gatcagcctt atgctaata aggtacagtt gtggtcactg 3000
aaagagtaat acagcctcat ggggtggat cgaatcctc ggaaggcact cagcatcttc 3060
aagatgtacc ttacgtcatg gtgagggaaa gagagagctt ccttgccccc agctcagggtg 3120
tgcagcctac tctggccatg cctaataatg cagtaggaca gaatgtgaca gtgacagaaa 3180
gagttctagc acctgcttcc actctgcaat ccagttacca gattccact gaaaattcta 3240
tgacggctag gaacaccacg gtgtctggag ctggagtccc tggccctctg ccagattttg 3300
gttttagagga atctgggtcat tctaattcta ccataaccac atcttccacc agagtcacca 3360
agcatagcac tgtacagcat tcttactcct aaacagcagt cagccacaaa ctgaccaga 3420
gtttaattag cagtgactaa tttcatgttt ccaatgtacc tgatttttca tgagccttac 3480
agacacacag agacacatac acattgatct taaaattttt ctcagtcact gatatgcaaa 3540
ggaccacact gtctctgctt ccaggagtat tttagaaatg ttccacaatt tactgaagac 3600
atagagatga tgcctgtgct taggtgcctt ttagcaagct atgcaaaca tcttgataaa 3660
acaagataca tagagagtca atctggcttc tgagaattta ccaagtgaac agagtaccta 3720
gttcatcagc cgtccagtaa agcaaccacg gaaactgact gggctctctt gcctaccgta 3780
ttaacattaa acattgatgt tctgtattct gtactttact gcaccagca gactttcaac 3840
aactcattga cccaaagtgc tgggattaca ggcgtgagcc actgcgccc ggcacattca 3900
gttcttatca aagaaataac ccagacttaa tcttgaatga tacgattatg cccaatatta 3960
agtaaaaaat ataagaaaag gttatcttaa atagatctta ggcaaaatac cagctgatga 4020
aggcatctga tgccttcacg tgttcagtca tctccaaaaa cagtaaaaaat aaccactttt 4080
tggtgggcaa tatgaaattt ttaaaggagt agaataccaa atgatagaaa cagactgcct 4140
gaattgagaa ttttgatttc ttaaagtgtg tttctttcta aattgctgtt ccttaatttg 4200
attaatttaa ttcattgatt atgattaaat ctgaggcaga tgagcttaca agtattgaaa 4260
taattactaa ttaatcaca atgtgaagtt atgcatgatg taaaaatac aaacattcta 4320
attaaggct ttgcaacaca aaaaaaa 4347

```

&lt;210&gt; 6

&lt;211&gt; 2116

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens





&lt;400&gt; 6

```

tagcgacctc tcgcagggaa agtcagcgtc ggccaaaagc ctccgggatc ggaatgagga 60
ggctgctgga gaagttgctt tctcctaaaa gggattatcc caggcgcaca cggtcattac 120
acgccgggga cgctcagtcg gtgcgggtac ccctgggcag ggccagcccc gcattccagg 180
ttctccatgt gcctagaaga cagtaatcga cggtatagca acagatctga ctgctaacat 240
gcgaaaccga tcagtagtag cagtagcacc agcaacagca gcacgaaaag caaaactaat 300
ctaaacggcc ctcagggctc aagcaggccc gacgaagact cgcccatccg gtcgccagaa 360
aactgggagt cccgcccgcg tttccggcac tgaaacgcga tcggccctgc ctggtaccgc 420
atcctcctct tgaccccacc tacactacga cgacggacgc tcgagaacgc ggcaccgcgc 480
cccgaggaa gtgcttccct gggcggaagc ttctgagcgt gatatagcgg aagtgccttc 540
tcttcgggtc tttctggtct cggccgcaga agcgagatga cgaagggaac gtcctcggtt 600
ggaaagcgtc gcaataagac gcacacgttg tgccgccgct gtggctctaa ggcctaccac 660
cttcagaagt cgacctgtgg caaatgtggc taccctgcc aagcgcaagag aaagtataac 720
tggagtgcc aggctaaaag acgaaatacc accggaactg gtcgaatgag gcacctaaaa 780
attgtatacc gcagattcag gcatggattc cgtgaaggaa caacaccta acccaagagg 840
gcagctgttg cagcatccag ttcattctaa gaatgtcaac gattagtcat gcaataaatg 900
ttctgggttt aaaaaataca tatctgggtt tggttaaggta tttttaatca attaggcttg 960
tagtatcagt gaaatactgt aggtttaggg actgggctag cttcatatca gatttacttg 1020
ttaagtgact gttttggaat gtttactttt ggactgggtt tgtaacacgg ttaaaggcaa 1080
tgagaaacaa gcagaattcc aggagtcctt gaagcagagg gcaactggaag acaatatagc 1140
agattaaaat agcacagctc atgtggcata ggtgggtatt ttagatgttt gagtaaat 1200
gaaagagtat gatgttttaa ttacctttag caacatgttc atctgctatg ctgtcatgac 1260
tagggggatg attattagtc acatagagct tgggagtacc actggaaacg tatgggtagg 1320
agtttaggtg gcttctgttt ttcaaaagat gatcttatcc tagtatctgt aatgctcact 1380
tggcacacct gacttgtggg ctgtgtgtaa ggtggctagc taagtgaata aagcctgcta 1440
ggtgtgagtc aacttaagaa tatgtaaata ggtttgagaa aaagtagggc ttgggtgcaa 1500
gtaaagattg agcaggaaat aaaggaaaat caagtataat ccctgagatt tntagattaa 1560
aggcaatgat gtgggactac ttggtcgaat ttttttagcc ctcaacttgg taattgggtg 1620
tttctgtgtt aaagcactga aacttgcgtg cgtgccttcc tagttttcgt ggtttattga 1680
cagggttggg ggtttttttt gtttttttaa aatgaaggga caaagtcaac tggactgctg 1740
agtgagaggg caggggcagt tgaagggaac atgaattgct ggaacagcta cataaaatag 1800
tgatgtagcc aagtcatgct atttaaatta taattctcca ctgtgtttag aataacatct 1860
gaggttctta acctggcctt ggaagggtat cacttttact tgtaacctgg aatggcttta 1920
taatgtgcta gctaattgct actctcatct tgtattttta ctctaattt acccttcagg 1980
tctcagcttc agaacattca cttataaaga aacctgctg attaaatctc tcttgggctt 2040
cctcccgaat tgtgagacta tactttaaag atgtatgggt agagtccaat tgccattgcc 2100
tttcttgggt acagat 2116

```

&lt;210&gt; 7

&lt;211&gt; 4474

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; unsure

&lt;222&gt; (9)



&lt;400&gt; 7

cggcccggn c ggggggggcaa gatggcgggc gacgtagggg ttcgtggccg gtacgagctg 60  
 ccgccttgct ccggcccagg ctggctcctc agcctttccg ccttgctgag tgtggcggca 120  
 cgaggggctc tggccaccac gcaactgggtc gtcacggagg acgggaaaat ccagcagcag 180  
 gtggattcac caatgaactt gaagcatcct catgacctag tcatattaat gagacaagaa 240  
 gcaacagtta actacctcaa agaattagag aaacaattag ttgctcaaaa aattcacata 300  
 gaagagaatg aggacagaga cacaggactg gaacagagac ataataaaga agaccagac 360  
 tgcatacaag ccaaggtgcc cttaggggac ctggatctat atgatggcac atacataact 420  
 ttggagagca aagacatcag tcctgaagat tatatagaca cagaatctcc tgtccctcca 480  
 gaccagagc aacctgattg tactaaaatt ctagaacttc catatagtat acatgctttt 540  
 cagcacttga gaggtgtaca ggagagagtt aatctttctg cacctctgct acctaaagaa 600  
 gacccaatct tcacataatt atctaaacgg ttaggaagga gtatagatga cataggtcac 660  
 ctcatctcatg aaggcctaca gaagaacact tcctcgtagg tactgtataa catggcttca 720  
 ttttactgga gaattaagaa tgagccatat caggtagtag aatgtgccat gcgagcactt 780  
 cacttctctt ccaggcaca taaagacatt gccctgggtc acctggcaaa cgttctacac 840  
 agagcacact tctctgctga tgctgctgtc gtggtccatg cagctctgga tgacagtga 900  
 ttcttcacca gctattacac tttggggaat atatatgcaa tgcttgggga atataaccac 960  
 tcagtgtctt gttatgacca cgctttgcag gccagacctg ggtttgagca agctataaag 1020  
 aggaagcatg ctgtcctatg tcagcaaaaa ctggagcaga aattggaggc tcagcataga 1080  
 tctctccagc gaacactgaa tgagttaaaa gagtatcaaa agcagcatga ccactacctg 1140  
 agacagcagg aaatcctaga aaacataaa ctgattcagg aggagcaaat cttagaagaa 1200  
 atcattcatg agactcagat ggcaaaaag gcaaatag gaaatcatca gatatgccga 1260  
 ctggcacaac agcagcatag ttacattgct cagtgggacc agcctgtacg ctatcatcgt 1320  
 ggagatatct ttgaaaatgt ggactatggt cagtttggtg aggattcatc aacctccagt 1380  
 atgatgtctg tgaactttga tgttcaatca aatcagagt ataatcaatga ttcgggtcaag 1440  
 tcttctcccg tagcccatc tattctctgg atttggggca gggactctga tgcataatag 1500  
 gacaaacagc atattctatg gcctaaaaga gcagattgta cagaaagcta ccctagagtc 1560  
 cctgttggtg gggaattgcc aacgtatttt ctgcctccgg aaaacaaagg actcaggatc 1620  
 cacgaactca gcagtgatga ttattctaca gaagaagagg cccaaacccc tgactgttcc 1680  
 ataactgact tcagaaaaag ccacactctg tcctacttag tcaaagaatt agaggttcgc 1740  
 atggatctga aagccaaaat gccagatgac catgcacgaa aaattttgc tccccgtatt 1800  
 aataactata ctatcccaga agaagaaatt gggctcttct tatttcatgc tattaataag 1860  
 ccaaagtctc ctatctggt cactactcaat gaagctggac tatactggag agcagtagga 1920  
 aatagcactt ttgctattgc ctgtcttcag agggctttga atttagctcc acttcaatac 1980  
 caagatgttc ctcttgctcaa cttggccaac cttttgattc attacggcct tcatcttgat 2040  
 gccactaagc tgctacttca agctttggcc atcaatagct ctgagcctct gacctttttg 2100  
 agcctgggaa atgcttacct tgctctgaag aatatcagt gggcacttga ggcctttaga 2160  
 caggccttga aattaaccac caaatgtcca gagtgtgaaa acagcctgaa gttgatccgc 2220  
 tgtatgcagt tttatccttt tctgtacaac atcacttctt ctgtttgcag tggtagcgtg 2280  
 gttgaggaga gcaatgggtc tgatgagatg gagaattcag atgaaaccaa aatgtcagaa 2340  
 gaaatactgg ctttgggtgga tgaatttcaa caggcatggc ctttgggaagg ctttgggggt 2400  
 gcactagaga tgaaagggcg gcgtctagac ttacaaggaa tacgggtgct gaagaaaggt 2460  
 cccagagatg gagtggccag aagctcttgc tatggagact gcagaagtga agatgatgaa 2520  
 gcaacagaat ggattacatt ccagggtcaa cgtgtaaaga aacccaaagg agatcataag 2580  
 aaaactcctg ggaaaaaagt agaaacaggt cagatagaaa atggacatcg ttaccaagca 2640  
 aacctagaga tcaactggcc caaggtggca tctcctgggc cacaaggaaa aaacgtgac 2700  
 taccagcgtc tgggatggcc cagcccgac gaatgcctca aactccgctg ggtagagctg 2760



WO 02/08278

```

actgccatcg tgagtacctg gcttgagtt tcttcaaaaa acattgacat cacagaacac 2820
atagattttg ccacccctat acagcagcca gcaatggagc ctctttgcaa tggcaatctc 2880
cccacgagta tgcataccct ggaccacttg catgggggtt ccaaccgagc cagcctgcac 2940
tacacagggg agagtcagtt aacagaggta ttacaaaate tcggcaaaga ccaatatcca 3000
caacagtcgc ttgaacagat tggcaccoga attgccaaaag ttttggaana gaaccagacg 3060
tcttggttcc tctccagcat ggcagccctc tactggaggg tgaaaggcca aggaaagaag 3120
gcaatcgact gcctccgcca ggctctgcac tatgcgccac accagatgaa ggatgtgccc 3180
ctgattagcc tggccaacat cttgcacaat gccaagctct ggaatgacgc cgtcatagta 3240
gccaccatgg cagtagagat cgcaccacac tttgctgtga accacttcac tctgggcaat 3300
gtctacgtgg caatggaaga atttgaaaaa gcactgggtg ggtatgaatc cacattgaag 3360
cttcagcccc agtttgtccc agccaagaac cgaatccaga ccatccagtg tcaactaatg 3420
ctgaagaagg gacggcgctc tccttagtgc acttcttctt tctctcttct tctttactca 3480
tgctctaaaa aaaaaaata agaaaagaaa ccaatcattg tcagtatcta ctattaatga 3540
tgtgtgtgaa aataactaag acttataaca ggacttttac atatgtggga attggtttgt 3600
ttttgttttt acgttttctc tttcccccaa ccaacctcag aagaggcacc ttcagaaaca 3660
cacatttctt aaaaggaaa tgcagcttca agatattgtg taaatactga gccaagacat 3720
ttctggagct gtgctctgtc tccaaaaacc tcaatgcctt tagggctttt ctcagtggtc 3780
cagctagcct tctcttttga ggaggatgaa gccgcattgc acattctctg cttcctgtcg 3840
tagcctctgt tgtcaatgga aatgcggaag cccatctggt gcccgtcagt gagaagcaac 3900
gttctgcgct ctctccggtg gacctccatg ctgtccccag tcttgtccat tccatgctgc 3960
tgtgttacaa actctcagag gtagtttgca ggggaggaag ggggaatatga ttttaaaaac 4020
aaaatattta caacaacaaa aattcttagg atcacctgac ctttgtaatg ttatttatgt 4080
tggggaggga ggggggctga gaaggggaaa tcagcagtg gcaacatctt tataatttgt 4140
actttaatta caaatcacia ggaaaccaat aagttgaaat cctatataac aggtttatat 4200
atatagaata tgtatatttg aagccctcta cagactgagt ctatgtttta ctaattcttt 4260
gttactgtg ttacccatct tggaaataagt tgtgaatgtc agctccctct ctctgaggcc 4320
tccagactta gtcctcagg agggtaatga gccaagggtg agtgtttcca tacaatgctt 4380
ttaccttga tcccaggaga atcagaaact ccaacatttt ggaatcttca agggcacata 4440
ctgagaaaaa aaataaaatt gtttatgagc aaaa 4474

```

&lt;210&gt; 8

&lt;211&gt; 777

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; unsure

&lt;222&gt; (269) .. (439)

&lt;220&gt;

&lt;221&gt; unsure

&lt;222&gt; (715) .. (716)

&lt;220&gt;

&lt;221&gt; unsure

&lt;222&gt; (743)



<220>  
 <221> unsure  
 <222> (750)

<220>  
 <221> unsure  
 <222> (753)

<400> 8  
 aaataaataa ataaaatttta gattaatttg ctgttatatt tttatataaa ctatgacata 60  
 agtataaaca aaaaaataga ataagtaaat aaataaataa aatttagatt aatttgctgt 120  
 tacattttta tataagctat gtttatgaca gactttccta taatattctt atcataatgt 180  
 tcttgcaactt gaaagaatgt gcattctgca gttgtgtgca ggtgttatgt gtatttcaac 240  
 tgggtcaagt ttgttaatca gggtggcnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 300  
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 360  
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 420  
 nnnnnnnnnn nnnnnnnnna cgttttctgg tagagtgaac cctttattat gtgaaatggc 480  
 tttctctacc actcacagtg gtattttttt cactttcccc cttatttttg aattttcagt 540  
 gttcttataa tgatgtcatg tctcttgaat agagtgtgtt taatcaagtc tgtcaatttt 600  
 tgtcttttgt gtatttctgt aatttacatt tattgtaatt actgatatag ttgtgttcat 660  
 aaataccatc ttactatttg tttcccatgt gtcttaccga tttttgtttt tattnnnttc 720  
 ttttctgttg cctttccttt ccngattaan ttnttaattc ccgttttccc ccccta 777

<210> 9  
 <211> 3195  
 <212> DNA  
 <213> Homo sapiens

<400> 9  
 ctctctttta gtgtcactgt caatggcgct acatggactt tgtaataacc ctttgaggca 60  
 catagctggg tgccatgtag aacatgtatc tgttacgata agtgtgtgcc caagaaatca 120  
 gaagaatgga ctttaatctc attttagaaa gtatgatatt aaatgattta cccaagccat 180  
 acttcagggt aatgacacga tagaagctag tacctgtgtc tctcaaattt ttctaacact 240  
 ttttatcttc cgatcagggt tgtgcagggt aaactcagt gagaggaaga tatatatccc 300  
 cttaaataaa acagctccct gtgttcgcct gctcaacgcc actcatcaga ttggctgcc 360  
 gtcttcaatt agtgagaca cagggttat ccacgtagta gagaaagagg aggacctaca 420  
 gtgggtattg actgatggc ccaaccccc ttacatggtt ctgctggaga gcaagcattt 480  
 taccagggat ttaatggaga agctgaaagg gagaaccagc cgaattgctg gtcttgcat 540  
 gtccttgacc aagcccagtc ctgcctcagg cttctctcct agtgtacagt gcccaaatga 600  
 tgggttttgt gtttactcca attcctatgg gccagagttt gctcactgca gagaaataca 660  
 gtggaattcg ctgggcaatg gtttggttta tgaagacttt agtttcccc tctttcttct 720  
 tgaagatgaa aatgaaacca aagtcacaa gcagtgtat caagatcaca acctgagtca 780  
 gaatggctca gcaccaacct tcccactatg tgccatgcag ctcttttcac acatgcatgc 840  
 tgtcatcagc actgccacct gcatgaggcg cagctccatc caaagcacct tcagcatcaa 900  
 cccagaaatc gtctgtgacc cctgtctga ttacaatgtg tggagcatgc taaagcctat 960  
 aaatacaact gggacattaa agcctgacga cagggtgtg gttgtgcc cccggctgga 1020  
 tagtcgttcc tttttctgga atgtggcccc aggggtgaa agcgagtggt ctctcttct 1080





caccagctg gctgctgctg aagctttgca aaaggcacct gatgtgacca ccctgccccg 1140  
 caatgtcatg tttgtcttct ttcaagggga aacttttgac tacattggca gctcgaggat 1200  
 ggtctacgat atggagaagg gcaagtttcc cgtgcagtta gagaatgttg actcatttgt 1260  
 ggagctggga caggtggcct taagaacttc attagagctt tggatgcaca cagatcctgt 1320  
 ttctcagaaa aatgagtctg tacggaacca ggtggaggat ctccctggcca cattggagaa 1380  
 gagtgggtgct ggtgtccctg ctgtcatcct caggaggcca aatcagtccc agcctctccc 1440  
 accatcttcc ctgcagcgat ttcttcgagc tcgaaacatc tctggcgttg ttctggctga 1500  
 ccactctggt gccttccata acaaatatta ccagagtatt tacgacactg ctgagaacat 1560  
 taatgtgagc tatcccgaaat ggctgagccc tgaagaggac ctgaactttg taacagacac 1620  
 tgccaaggcc ctggcagatg tggccacggg gctgggacgt gctctgtatg agcttgacagg 1680  
 aggaaccaac ttcagcgaca cagttcaggc tgatcccaa acggttaccg gcctgtctta 1740  
 tgggttcctg attaaagcca acaactcatg gttccagtct atcctcaggc aggacctaag 1800  
 gtctacttg ggtgacgggc ctcttcaaca ttacatcgct gtctccagcc ccaccaacac 1860  
 cacttatgtt gtacagtatg ccttggcaaa tttgactggc acagtgggtca acctcaccg 1920  
 agagcagtgc caggatccaa gtaaagtccc aagtgaagac aaggatctgt atgagtactc 1980  
 atgggtccag ggccctttgc attctaata gaacggaccga ctccccgggt gtgtgcgttc 2040  
 tactgcacga ttagccaggg ccttgtctcc tgcccttgaa ctgagtcagt ggagctctac 2100  
 tgaatactct acatggactg agagccgctg gaaagatata cgtgcccggga tatttctcat 2160  
 cgccagcaaa gagcttgagt tgatcaccct gacagtgggc ttcggcatcc tcatcttctc 2220  
 cctcatcgtc acctactgca tcaatgcaa agctgatgtc cttttcattg ctccccggga 2280  
 gccaggagct gtgtcatact gaggaggacc ccagcttttc ttgccagctc agcagttcac 2340  
 ttcctagagc atctgtccca ctgggacaca accactaatt tgtcactgga acctccctgg 2400  
 gcctgtctca gattgggatt aacataaaa agtggaacta tccaaaagag acaggggagaa 2460  
 ataaataaat tgccctccct cctccgctcc cctttcccat cacccttcc ccatttctc 2520  
 ttccttctct actcatgcca gattttggga ttacaaatag aagcttcttg ctctgttta 2580  
 actccctagt taccaccct aatttgccct tcaggaccct tctactttt ccttctgccc 2640  
 ctgtacctct ctctgtcct cacccccacc cctgtacca gccaccttcc tgactgggaa 2700  
 ggacataaaa ggtttaatgt cagggtcaaa ctacattgag cccctgagga caggggcatc 2760  
 tctgggctga gcctactgtc tccttccac tgcctttct ccaggccctc agatggcaca 2820  
 ttaggggtgg cgtgctgcgg gtgggtatcc cacctccagc ccacagtgtc cagttgtact 2880  
 ttttattaag ctgtaatatc tatttttgtt tttgtcttt tcctttatc tttttgtaaa 2940  
 tatatatata atgagtttca ttaaaataga ttatccaca cgacttgta tgctagtatt 3000  
 tcttcccagg ccacctgtt cagcgagcct agactggaag tcatgaagtt atcttttatg 3060  
 ctatcatctt gggctccaga ggaccaagg agtaaggctc tgtcaaaaac agttgaagtc 3120  
 ctttcaaatt gcagagcctg gttctcctct tgtaagaaca atgttaacat agtttcttct 3180  
 cactttgtaa tgaac 3195

&lt;210&gt; 10

&lt;211&gt; 949

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 10

agcctccacc ctggcgatgg ctccctgggc ctactttctc tctcaaactg gctttttctc 60  
 attcctttga ctccgccaga cttcctcgcc cccatgacct ggtgttgtgt ctgatcccc 120  
 caacattcct ggctgccccaa tgtggggcaa tgaagacccc agtgaaggaa tgctagagt 180  
 tgtgaaagtg gaggacgcat cgtcaaagga cacctgagga cgtctcaaag aagctcggcg 240



WO 02/08278

ggagagctga ggcgtcggaa gaaccaagaa tcattctctt tgaaaaatcg attcatcaaa 300  
 tgaatcttca gccacaact gttcaagaag gattcaaata tcacagggtc caagaagtaa 360  
 agctttggag gtcacaaaat tagcaataga agctgggttc cgccatatag attctgctca 420  
 ttatataaat aatgaggagc aggttggaact ggccatccga agcaagattg cagatggcag 480  
 tgtgaagaga gaagacatat tctacacttc aaagcttggt ccacttttca tgcaccagag 540  
 ttggtccgac cagccttga aaactcactg aaaaaagctc aattggacta tgttgacctc 600  
 tatcttattc attctccaat gtctctaaag ccagggtgagg aactttcacc aacagatgaa 660  
 caagtggcaa aagtaatat tgacatagtg gatctctgta ccaccctggg agggcatgga 720  
 gaaagtgtaa ggatggcagg aattggggca agtccattgg ggtgtcacac ttcaaccgcg 780  
 aggcgctggg gatgagtctc aaaaagcagg aatccagtta aagcggctctg caccggtgga 840  
 gtgtccgatt taccgggtaa tgctgattcg gcagccagaa atgttggtgc aaagctgggg 900  
 cccacaaaat ggggcccccc ccgctgtggg ccctggttga aaaacctg 949

&lt;210&gt; 11

&lt;211&gt; 14917

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 11

gctcgagatc cattgtgctc taaagggtgaa aagacctaga tggtagaagg caagcagtgg 60  
 aatgtgtctg ggtatccgag aaatatacag aaagcactta agagaactta atcatttttt 120  
 tctccctttc cttagattga atagggaaaa cctgctttct gcaaacaaact gaaaaagctg 180  
 catttagaaa ctgcttcttt ggccctcatc gagaagctgg aacttgaatt gttaggcccc 240  
 ttatgggaca agctctcaac tgctgatcac ccagtgattg acaccatggc cagcaagagg 300  
 aaatccacca caccatgcat gatcccagt aagactgtgg tgttgcaaga tgccagcatg 360  
 gaggcccgag ccgctgagac cttgcctgaa ggaccccagc aggatctgcc cccagaagca 420  
 tctgctgcca gcagtgaggc agcacagaac cccagcagta ctgatggctc tacactggcc 480  
 aatgggcctc ggagcacttt agatggctat ttatatctct gtaaatactg cgatttcaga 540  
 tccccatgaca tgaccaatt tgtgggacat atgaactcag agcacacaga ctttaataaa 600  
 gacccaacct ttgtatgcag tgggtgcagt tttctggcaa aaacctctga ggggcttttc 660  
 ttgcacaatg ccacatgtca ctccggggaa gccagctttg tgtggaacgt ggccaagcca 720  
 gacaatcatg tgggtgtgga gcagagcatc cctgagagca ccagcactcc tgacctagcg 780  
 ggtgagccca gtgctgaagg ggctgatgga caggcagaaa tcattcattac caaaactcca 840  
 atcatgaaga taatgaaagg caaagctgaa gccaaaaaaa ttcatacact caaggagaat 900  
 gtccctagcc agcctgtggg tgaggcctta ccaaagctgt cgactggaga aatggagggtg 960  
 agagaggggg accattcctt catcaatggg gcagttccag tcagccaggc atctgccagc 1020  
 tctgcaaaaa acccccatgc cgccaacggg ccctgatag gaacagtgc agttttgcca 1080  
 gctggcatag cacagttcct ctccctccag cagcagcccc cagtgcattg ccaacaccat 1140  
 gtccaccagc cactgcccac ggccaaggcc cttcccaaag tgatgatccc cctgagcagc 1200  
 attccaacgt acaatgcagc catggactct aacagcttcc tgaagaactc cttccacaag 1260  
 ttcccctacc ccaccaaagc cgagctctgc tatttgactg tggtgaccaa gtatccagaa 1320  
 gaacagctca agatctggtt cacagcccaa aggctgaagc aggggatcag ctggtcccct 1380  
 gaggagattg aggatgcccg gaaaaagatg ttcaatacag tcattccagtc tgtgcctcag 1440  
 cccacaatta cgggttctaaa taccctcctc gtgcgagcag ctggcaatgt ccagcatctc 1500  
 atccaggccg ctcttccagg tcacgttggtg gggcagccag aggggtacagg agggggactt 1560  
 ctggtcactc agccattgat ggccaatggg ttgcaagcaa caagttcccc tctccccctc 1620  
 acggtgacat ccgtcccaa gcagccaggt gtggcaccca ttaacactgt gtgttcaaat 1680



acaacgtcag ctgtgaaggt ggtcaatgcg gccagtcgc tcctcacggc ctgccccagc 1740  
 ataacctccc aagccttcct tgatgctagc atctacaaaa ataagaaatc tcatgaacag 1800  
 ctgtcagctc tgaaagggag cttctgtcgg aaccagttcc cagggcagag cgaagttgaa 1860  
 catctcacia aagtgcggg cctcagttacc agagaggtgc ggaaatggtt cagtgtatcgt 1920  
 agataccact gccggaactt gaagggctcc agagcgatga tacctggaga tcacagttcc 1980  
 atcatcattg actctgtgcc agaggtgtcc ttctcccat cgtccaaggt ccctgaggta 2040  
 acctgcattc cgacaacagc cacactagca accaccctt ctgccaaacg acaatcttgg 2100  
 caccagactc ctgacttcac accaaccaaa tacaaggaga gagccctga gcagctcaga 2160  
 gccctggaga gcagttttgc acaaaaccct cttcctcttg atgaggaaact ggaccgcctg 2220  
 agaagtgaag ccaaaatgac ccgacgagaa attgatagct ggttttcaga gagacggaaa 2280  
 aaagtgaatg ctgaggagac caagaaggct gaggagaatg cctctcagga ggaagaggag 2340  
 gctgctgagg atgaggggtg agaagaggat ttggccagtg agctaagggt ctctggtgaa 2400  
 aatggctctc tggaaatgcc cagcagccat atcttggcag agcgcaaagt cagccccatt 2460  
 aaaatcaacc tgaagaacct gagggctact gaagccaatg gcaggaacga gattccaggg 2520  
 ctgggtgcct gtgaccctga ggatgatgag tcaaacaaac tggcagagca gctcccaggc 2580  
 aaagtgcgct gcaaaaagac tgcccagcag cggcacttgc tgcggcagct ctttgtccag 2640  
 acacagtggc caagcaacca ggactatgac tccatcatgg cccagacggg tctgccacgg 2700  
 ccagaggtgg tgcgctgggt tggagatagc aggtacgcac tgaagaacgg ccaactcaaa 2760  
 tgggtacgaag actataagcg aggcaacttc ccaccagggc tactgggtcat tgccctggc 2820  
 aaccgggagc tcctgcagga ctattacatg acacacaaga tgctgtatga agaggacctg 2880  
 cagaacctct gtgacaagac ccagatgagc tcccagcagg tcaagcagtg gtttgctgag 2940  
 aaaatggggg aggagaccag agcgtggca gacacaggca gtgaggacca gggccctggt 3000  
 actggtgagc tcacagcagt tcacaaaggg atgggtgaca cctattcaga ggtgtctgag 3060  
 aacagtgcgt cgtgggagcc ccgtgtccct gagggcagct cagagccctt tgacacatcg 3120  
 agtccccagg ctggacgtca gctcgaaaca gactgaattt gatctgatta atgtgaagga 3180  
 ctggccagtc tgggaaaccg cctgccacgt ggaagagcca aaccgcactc tctgctgcca 3240  
 catgccgttc ccatgcccgg ctgctgggca cctgggagag cttccagaat cctcgcagac 3300  
 agcccagagc ctgccgtac cctcggcctg cccaccacca agcaagcagc aagcaagatg 3360  
 gggttctcat cagttcttcc tcccacaatg taggaccttt cctttacctt ccaatggata 3420  
 aaatagttca gagttcatag tcatattcat agacacagaa tcaagctttt aacatataca 3480  
 tccacctcta tatgttaaata aaaacatcag attatcaaca ctgtcattac gtagaaactt 3540  
 tggttagcca agcagtgcac gtgcagttac gtcactctca aaaatgacct gtgtctgttc 3600  
 tctggggatt gctgggtcac aggtgcccct caccctccac agtcaggcag ggaagttata 3660  
 ggcacaaagc tacgtctgga acccctttgt gcccccttg tgttcctcaa ggaagcagta 3720  
 cctttgaaga gatctctgct gcattaagtg atgaccggct acgtttcatg tcaggcttgc 3780  
 tttgccttgt gggctactca gtgcagaacc tgctgtaacc ctcagttcaa aaaatggact 3840  
 ggcaatgtga ttagcgttgg atgctttacc attctcttta gttgttaccg taattctgct 3900  
 ttttcatggg agtttgaatc atggaccata acttttcagt tatcagatca actaaagaaa 3960  
 catttgttgt taagcctaata gtgctgacct atgtgcctgc attttttttt taatctagac 4020  
 atgtttggag tgagagaaaag atggaaaaaa gacatggggt agggacgtaa gtggaaatct 4080  
 atagccacag cctgaagctt tgaccactgc ggttttcaga gccctttctc cacactcatt 4140  
 tcagagcctc ctatgggttg ggaaggaata acacactggc ccattagtaa ggggtgaaggc 4200  
 tggagggatt tgttgacttc ttggaattat cagaggtagg gtggtcttta gcacaaagac 4260  
 ttgcatgcag agatccctgg cagaacaccc agagtgcctg tggctccac cccagggtct 4320  
 ggccgggtgt ctggatgcat gcccaagggt gctgggcctc actggtcctt gtgaggatgc 4380  
 ttttaaagtt ttatatttat gtcccaaac ttggaaacaa gaactctact ttagcctaac 4440  
 cctcatgtcc ttttttgaat tgagaaaatt acaggaaatg gtgcctttga aaattagaaa 4500  
 acttgcttac agagctgttc taaatggtaa atcctcaatt tcccaagac cggttgctct 4560



gagagtagct ggtaaagagg ggcgaactaa agacctgtcc acctgtagct ccgctcattt 4620  
 cttagaaacc acctgcttcc cagagtgcc agccacaagt accaggcttc gtgggcacag 4680  
 acacctcctg ggctgggcag agtgacagtg ctagaagacc ccagagagag ggcaagggct 4740  
 ttgggcaaga agcactggtg gtgttttagg gacctgcctt cccccctacc cagggaactg 4800  
 gacctggcag ggcaccgtgc tcatgtggct ccaaggacaa gcatggcggt ggcccccttct 4860  
 gccttccagg agaggctctg cttttgaaac caaaatcatg ttcttctaaa gtgtcatctt 4920  
 ctgccccccc tgtccaata gggaccacat cttatttgtc tcaaacaggg acttgtgagt 4980  
 acttggcaag ttttgcagcc tatttttgta ttcttaattt ggggagtaaa gatgtttggt 5040  
 ctcaaaaacc tttgaggaat tgccaagaat ggcgagtgtat tgctttcctt cagagaacag 5100  
 acacttgga tttctccttt tagtgtttat atacgtgcag attaatattt atatatatat 5160  
 atacacacac acatatacag tataaatact catttgattc tcgtaaaacg cgcattctggc 5220  
 gtgtgcagtt gagaaacttg gtggcacatg ggtgttgggg gagtagcctg tgttggaggg 5280  
 acaccagtgc actaggcagc tggggcgggc caggctgaag ccattctccg gtgtctgaga 5340  
 aaccaccag tgccctacct ccagatcctg ctggcatcac ctccagagcc ctgcatgcac 5400  
 tggctgaaga gttggtctgt ggagaggatt ttcttggtta cttgtattca cgtttaattt 5460  
 acaacccaaa cagcaaaaca cagttggtgg acaagttcat gcaggaccta cagtgaccca 5520  
 gccatgggca ctagctcatc tttcaggtgg aaaagtacag tgctgcctgc cctggtatgt 5580  
 ttttcttata gatgttagcc ctgcccaca gccagggcta cactacaaaa ggcaagaatg 5640  
 cccatgtaag gagcccagca gtctggacag atccttcttc ctctgctgtt ggatgagagt 5700  
 gagtgagtat gctctggacc ttatccttga aagatgatca aaagcgatga tgaggaggc 5760  
 agtcatcacg cagggtgctta aaggacatt gtaggaggtta ctcaagggtt tgggggcaaa 5820  
 acctgaatc cagccagtcg tgcacagaga cacaccaca ctagcccagt ggcagtgggg 5880  
 gatgtgggat ggcagcagcc aggtatgtag cctgtcaca ggacagctca ctgtggtttt 5940  
 gcacactgcc taagggttaa attgtgtgtg tgccttcagt agaaggcatt tgtgggctgc 6000  
 agagttgaga gttgggtgag gttagtctct cctgaagaaa aagcctataa aaagtggcta 6060  
 atcttatccc ttttctctgt atgcagttgg actcgtcaga gatagtaaaa tcatctttta 6120  
 gtgttttttt gttgctgatg tctgtaccc atttgttttt tacatggggt tgtagatcga 6180  
 gttctcaaag gtgaaaccag atgatcattc tgataaagga aatttaaatt tgatacatat 6240  
 gctttgtata ttttgattac ttgttttctg ttttgactat aaaggagctt ttttattttg 6300  
 ggaggggagg agtgtcattt ttgagaatct tgggttcctg aaaaagaacg ccctagtgtg 6360  
 atggcttgcc agggccttgg ggtttggtag tgattgtaca acttaaagct cctttctctt 6420  
 ggctgagtga cagggtgctg ttcaggtgga ccaaagcacc ttgacacaag gactccacac 6480  
 tgtgctctct agtagcaca ggaggaagtt ggacagaaca ttgggtagtg ccttgcgggc 6540  
 tcacacatgt actagtggtc tcatctccag ctagccttgg gagggcgtcc caccaggaaa 6600  
 tctctctatt ccgtagcctg agatgtgccc ttgtgggttt tatcctgctc agtcagtggc 6660  
 ctaggggcag gtctgtgtt ctctccctct ctctcctggc tctgggacat ctgtccctgg 6720  
 ctgcctttca tggaggaagg aactggcct ttctgggttg atgctgtgtg gatgctctct 6780  
 gcttgagcct cgtggtctct tgtctctttt tggaccaata tctcagatt ggtgcagctt 6840  
 tttcagttca gatatcacac ccaagtggga taaaggcaac ttgcaggaga ggagagccag 6900  
 ccaagaagaa aatttttaaa ccaaacctcg tttaggattt tcctaaagtc atcttctctt 6960  
 ttttcttgct cagagtttac ctgggagatt tcaccagttt gactcaccat ttgcagatgt 7020  
 gcttttgtat taaatttaaa ttttcacata tcacatccat tctcaaggta gttatatgct 7080  
 ggagaagaaa aatcctctag acacatgaag gccacatag tcaagtcttc cagggcaaaag 7140  
 ccagcagccc acccaggtca ggtagccagc agggctcagt tccccctact ccagacacgg 7200  
 acctcctctc tcagggtctc ttgaccagc ttcttctct ccttttacct gagagcacag 7260  
 acctctctca gccagcctgc ccagaccagc gggggctact cccatgtagt ttggggagca 7320  
 cttgatctca gaaaagctcc attgtctgag caaatgggca gttgtggagc tcaagccttt 7380  
 ctctgtgct caagtccctt cccaagcaa ggcttcaacc tcatctaccc accatgtagt 7440





tttctctggc catttaagtg gggcggcagg gacatgggtg ggccatgcc aaccagggtc 7500  
 ggtgaggcaa ccagttttga ttttgacaga gtggctggag gaaaagtggc aatcaagggtg 7560  
 ctgcttggtt tgctctgagt gcaaatggaa ccaacagggtt tctgctgcaa tctgtgtgtt 7620  
 cccagtgcc aagtcacacca ggaggggtgg ggcagggtta accaagtggc ctctgaactc 7680  
 accgagcgtc tgcacttggt tgtgaagtta atgggagtag agagagcgtc tggccttgga 7740  
 gaggggttga gagcctcctt tttgggttct cattcctgag ctcttgctg cccacaaatc 7800  
 tgacctcttt gaatggggac gcagtccttc aacagagaag tttctatggc aaagaagtgt 7860  
 ctatttagct ctagatccag cagagtcact cattctaact gccctgaagt ctagagcagg 7920  
 ggagggaaac cagaggctgg ggatgagact aggcagacc tggttaccat atggacaagg 7980  
 acaggggaaa gcacccctt cctcaatttc tgaaagtctt atctttgggt tcgcaggact 8040  
 ttgaggatga taaagaacat ataggtacta gcttgtgtgt gctgggtcaa agcttcaca 8100  
 gccctgagaa tttggcttct gtggctgctc tggcagctga gcgaaggag gaaggcagcc 8160  
 gctctgggtg ggactctagg caccttcctt gctgtccact tggataggcg gtgagccca 8220  
 gggtagctag aggagcctga gcatttacct gccattagt cctcttcctt caggagactg 8280  
 gcttgaaacg tgtgtcatg tgcgcgtgca cacacacaca tgagcacctg tatgtgttaa 8340  
 tgaatagttt ttcttggtta atgcttttta acttctgttc ctttccgtaa gtggatgatt 8400  
 caaaattaac gtgacttggt tgggcgcagt ggctcacacc tgtaatccca gcactttggg 8460  
 aggccaaagg caccagata aacagccaca ggtcagaagg ctgctgagt cccctggaag 8520  
 cagaatagct gggcaatggg tccttgactc tctgaaatct cctcatttac tgctgaaagg 8580  
 ggaaaatgac aagaatagtt tatcccagaa agcatttctc tacgttgctc atgttttga 8640  
 tgagtctgag agaggcgtgc tggtcacat gacaacagag acaggccccc actctgaggt 8700  
 gaagaaagct caggaggagg cccgcagca gcccgaggct gctgccgctg tgaccacccc 8760  
 tgtgaccctt gcaggccacg gccaccaga ggccaactcc aatgagaagc atccatccca 8820  
 gcaggacacg cggcctgctg aacagagcct agacatggag gagaaggact acagtgaggc 8880  
 cgatggcctt tcggagagga ccacgcccag caaggcccag aaatcgcccc agaagattgc 8940  
 caagaaatac aagagtgcc tctgccgggt cactctgctt gatgcctcg agtatgagt 9000  
 tgaggtggag aaacatggcc ggggccagggt gctgtttgac ctggtctgtg aacacctca 9060  
 cctcctagag aaggactact tcggcctgac cttctgtgat gctgacagcc agaagaactg 9120  
 gctggacccc tccaaggaga tcaagaagca gatccggagt gagggttctg ttgtgtttg 9180  
 ggaggtgggt agcccctgga attttgctt cacagtcaag ttctaccgc ctgacctgc 9240  
 ccagctgaca gaagacatca caagatacta cctgtgctg cagctgcggg cagacatcat 9300  
 caccggccgg ctgccatgt cctttgtcac gcatgcccta ctgggctcct acgctgtgca 9360  
 ggctgagctg ggtgactatg atgctgagga gcatgtgggc aactatgtca gcgagctccg 9420  
 cttcggccct aaccagacc gggagctgga ggagaggatc atggagctgc ataagacata 9480  
 tagggggatg accccgggag aagcagaaat ccaactctta gagaatgcca agaagcttct 9540  
 catgtacgga gtagacctgc accatgccaa ggactctgag ggcacgaca tcatgttagg 9600  
 cgtttgtgac aatggcctgc tcatctaccg ggaccggctg agaatcaacc gctttgcctg 9660  
 gcccaagatc ctcaagatct cctacaagag gagtaacttc tatatcaaga tccggcctgg 9720  
 ggagtatgag caatttgaga gcacaattgg ctttaagctc ccaaaccacc ggtcagccaa 9780  
 gagactgtgg aaggtctgca tcgagcatca tacattcttc cggctgggtg cccctgagcc 9840  
 cccacccaag ggcttctggt tgatgggtc caagttccgg tacagtggga ggaccaggc 9900  
 acagactcgc caggccagcg cctcattga ccggcctgca ccttctttg agcgttcttc 9960  
 cagcaaacgg tacaccatgt cccgcagcct tgatggagca gagttctccc gccagcctc 10020  
 ggtcagcgag aaccatgatg caggccctga cggtgacaag cgggatgagg atggcgagtc 10080  
 tggggggcaa cggtcagagg ctgaggagg agaggtcagg actccaacca agatcaagga 10140  
 gctaaagtct ttagacaagc cagaagatgt cttgtggaag caccaggcca gcatcaatga 10200  
 gctcaaaagg accctgaagg agcccaacag caaactcatc caccgggagc gagactggga 10260  
 acgggagcgc aggtgcctt cctccccgc cctccctcc cccaaggga cccctgagaa 10320



agccaatgag tcccagagga cccaggacat ctctcagcgg gacttggtac ctgagcctgg 10380  
 agcagccgca ggcttggaag tggtcactca gaaaagcctc gcagcatctc ctgaggggttc 10440  
 agagcattgg gtatttatag aaagagagta cactaggcca gaagagctcg gtctcctaaa 10500  
 agtgaccacc atgcagcagg aagaaaggca ggcaggcctt gctggatatcc ttgccaacgg 10560  
 cagactctcc aaggtagacg ttctgggtga caagttcaaa gtggaagtgg ccacagaaga 10620  
 aatgggtggga aacagaagag caaacaccca gcaacaagga aaaatgattg caagtcctga 10680  
 agactttgag tcagtggggg aggaaggccc ctggatcagg gaaagcccag gaggggctgc 10740  
 cctggcttcc ggccgcacat tggcagaaaa gctcctcgag ggctctgagc tcagggcaga 10800  
 caccagagag gcaaccatca ggaaccgctg catgtcagat ggtcagcccg agggccagac 10860  
 agagctgagg aaggggctgg aggagcctca cacttggtgg agaccactg ctccagggag 10920  
 caggccagca gagggtggacg tcctctctcc agcctccgac aagggaggac tccagtcgtt 10980  
 tctattggat ccagcccacg cagaagccag agctgagttg agcaatgaaa ctgatacttc 11040  
 ctttgcagag aggagcttct atttaaatta tgaagaaaa gactcagaag accaagtcct 11100  
 ccctccaccc ctggaggaga gaaaagggcg cctggatgcc cctcccggag gtgagcccag 11160  
 gccgacgctg aattccttag acctgagggt ttctgctgct gcttccagca ggagcaagga 11220  
 cgaagcccac atgacttccc caaagggaagg ggcaggggacc cccaagaacc atggaggacc 11280  
 tggtagacctg aagggatctc ccgcaggaca gacgtttgct gaaggctggg aagatgccc 11340  
 gtggggagtg gaaggagagt ttccccacct gacagccagc gcagcccag aggaaggag 11400  
 ccccgtagtg ggagatttgc tgggaaaggc tgaggaaagt cccacagagg aactgaagaa 11460  
 gcaccctcct cacagaggac agggcggtga tcccgacccc caggcctgcg cccttccctg 11520  
 ggccatccct ctgaatgtca ggaagccagt caaacagac agaggcaact tcccacccaa 11580  
 agagagggga gtggttccca cccagaaagg aggggctgag ctgaaggacc gcgaggcttc 11640  
 agcatttctt cacatggagg tgatcattcc cctgccagcc tcccctggtc attctgagga 11700  
 cctggcagct ctggaggaag cttctccaag cccaacctcc catgggtcag gggagccttc 11760  
 ggagctcagg gagcccttcc ttagacatgt ccattcttcg aaagccagcc cagagcccaa 11820  
 ggaccaagta gggtttgtgg tgtcccctgc cacaggaggt gagcgcaggc ctccctccat 11880  
 caccagcaga aagcccagag tagtccctga agaagctgag gggcgcatat ctctgggggt 11940  
 tgggttccct tcagggaagc gaaggagat gacctcttcc caggctgggg accaagaggg 12000  
 ctccctagaa gatattagca agacctcagt ggccaacaaa attcggatat ttgagaccca 12060  
 cggagctgaa actcgccgaa tgagtgggg tgaagcaagg tcccttccaa atgacgtatc 12120  
 ttcagaggca cccgtgggac aagcagagca gcagcggagt acgctctcag acctgggctt 12180  
 cgcccaactc cagccccag gggactttgc cagccccaaa gccacacatt ccacagtgat 12240  
 acctctgggt accagacact tcaggagga cacttctgca tccctaccag aagcacacac 12300  
 ggaactagag cccgtgtccc ccaattcagg ctgtgaaacc acgctggcag aagctactgg 12360  
 aactggggta actggccgca acaaatccgg agatgcggtc aggggaagaga agcgtccac 12420  
 caacttagca gccaacaccc ctgggaagg ggggcgcctg agatttgcca gccctcggg 12480  
 ccctcagaga gcagggtga gggagggtc cgaggagaaa gtcaaaccac cagtcctccg 12540  
 ggccccagag agtgacacag gcgatgagga ccaggaccag gagagggaca cgggtgttct 12600  
 gaaggacaac cacctggcca ttgagcgcaa gtgctccagc atcacggtca gctctacgtc 12660  
 tagcctggag gctgaggtgg acttcacggt cattggtgac taccatggca gcgccttcga 12720  
 agacttctcc cgcagcctgc ctgagctcga ccgggacaaa agcgactcg acactgaggg 12780  
 cctgctgttc tcccgggatc tcaacaagg ggcacccagc caggatgatg agtctggggg 12840  
 cattgaggac agcccggatc gaggggcctg ctccaccccg gatatgcccc agtttgagcc 12900  
 cgtgaaaaca gaaaccatga ctgtcagcag tctggccatt agaaagaaga ttgagccgga 12960  
 ggccgtactg cagaccagag tctccgctat ggataacacc caggttgatg ggagtgcctc 13020  
 agtggggagg gagtccatag caaccactcc ctccatcacc acggagacca tatcgaccac 13080  
 catggagaac agtctcaagt ccgggaagg ggcagctgcc atgatcccag gccacagac 13140  
 ggtggccacg gaaatccgtt ctctttctcc gatcatcggg aaagatgtcc tcaccagcac 13200



```

ctacggcgcc actgcgga aa ccctctcaac ctccaccacc acccatgtca ccaaaactgt 13260
gaaaggaggg ttttctgaga caaggatcga gaagcgaatc atcattactg gggatgaaga 13320
tgtcgatcaa gaccaggccc tggctttggc catcaaggag gccaaactgc agcatcctga 13380
tatgctggta accaaagctg tcgtatacag agaaacagac ccatccccag aggagaggga 13440
caagaagcca cagaagctaa aacgagaaac taagaataac aaaaggcaac cttgcatccg 13500
gcggaacctg caggcacggg ccattcttcc cgcgaccag ggctctgccg gaccgcttcc 13560
cccgtcgctc cagtcagacg ccaaagctgg agaacttccg gtgcgtttcc gctgtaccgg 13620
aacgtggggc gaggcgctgt tcatcaaaga aaaagggttc ttttggtcac ccaccactgg 13680
ccccatggct gccgtgcaga tggatcctga gctagccaag cgcctcttct ttgaaggggc 13740
cactgtggtc atcctgaaca tgccaaggg aacagagttt gggattgact ataactcctg 13800
ggaggtcggg cccaagtcc gggcggtgaa gatgatccct ccaggcatcc acttcctcca 13860
ctacagctct gtggacaagg ctaatccgaa ggaagtaggc cctcgtatgg gtttcttcc 13920
tagcctgcac cagcgggggc tgacagtgtc gcgctggagc aactcaggg aagaggtaga 13980
cctgtcccca gcccagagt ctgaggtgga ggccatgagg gccaacctcc aggagctgga 14040
ccagttcctg gggccttacc catatgccac cctgaagaag tggatctcac tcaccaactt 14100
catcagcgaa gccacagtgg agaagctaca gcccgagaat cgacagatct gtgccttttc 14160
cgatgtgcta cctgtgctct ccatgaagca caccaaggac cgcgtggggc agaactctacc 14220
ccgctgtggc attgagtga aaagctacca agagggcctg gcccggtac cagagatgaa 14280
gcccagagcc gggacagaga tccgcttctc agagctgcc acgcagatgt tcccagaggg 14340
tgccacgcca gctgagataa ccaagcacag catggacctg agctatgcc tggagactgt 14400
gctcaacaag cagttcccca gcagcccca ggatgtgctt ggtgaactcc agtttgcttt 14460
tgtgtgcttc ctgctgggga atgtgtacga ggcatttgag cattggaagc ggctcctgaa 14520
cctcctgtgc cggtcagaag cagccatgat gaagcaccac accctctaca tcaacctcat 14580
ctccatcctg taccaccagc ttggtgagat ccccgctgac ttcttcgtag acattgtctc 14640
ccaagacaac ttctcacca gcaccttaca gaaaggattc ttgattgaca tcagtgaaga 14700
cttcacgag gtccgttttc tcagaggtag agaacgcttt tctaaagcca acctactgtt 14760
ggatgagtcc ctcatgtccg tggatgccac cctgagaaag aaagctgaaa agttccaagc 14820
tcacctgacc aagaagttcc ggtgggactt tgctgcggaa cctgaggact gtgccccggt 14880
ggtggtggag ctccctgagg gcacgagat gggctaa 14917

```

&lt;210&gt; 12

&lt;211&gt; 1823

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 12

```

cccacttccg gagacctcac acaagatggc ggcacccgag gaacacgatt ctccgaccga 60
agcgtcccag ccgattgtg gaagaggagg aaactaaaac atttaaagac ctgggtgtga 120
cagatgtgtt gtgtgaagct tgtgaccagt tgggatggac aaaaccacc aagatccaga 180
ttgaagctat tcctttggcc ttacaaggtc gtgatatcat tgggcttgca gaaactggct 240
ctggaaagac aggcgccctt gctttgcccc ttctaaacgc actgctggag acccgcagc 300
gtttgtttgc ctagttctt accccgactc gggagctggc ctttcagatc tcagagcagt 360
ttgaagccct ggggtcctct attggagtgc agagtgtgt gattgtagg ggaattgatt 420
caatgtctca atctttggcc cttgcaaaaa aaccacatat aataatagca actcctggct 480
gactgattga ccacttgga aatacgaaag gtttcaactt gagagctctc aaatacttgg 540
tcatggatga agccgaccga atactgaata tggattttga gacagagggt gacaagatcc 600
tcaaagtgat tcctcgagat cggaaaacat tcctcttctc tgccaccatg accaagaagg 660

```



```

ttcaaaaact tcagcgagca gctctgaaga atcctgtgaa atgtgccgtt tcctctaaat 720
accagacagt tgaaaaatta cagcaatatt atatttttat tccctctaaa ttcaaggata 780
cctacctggt ttatattcta aatgaattgg ctggaaactc ctttatgata ttctgcagca 840
cctgtaataa taccagaga acagctttgc tactgcgaaa tcttggttc actgccatcc 900
ccctccatgg acaaatgagt cagagtaagc gcctaggatc ccttaataag tttaaggcca 960
aggcccggtc cattcttcta gcaactgacg ttgccagccg aggtttggac atacctcatg 1020
tagatgtggt tgtcaacttt gacattccta cccattccaa ggattacatc catcgagtag 1080
gtcgaacagc tagagctggg cgctccgaa aggctattac ttttgtcaca cagtatgatg 1140
tggaactctt ccagcgcata gaacacttaa ttgggaagaa actaccaggt tttccaacac 1200
aggatgatga gggttatgatg ctgacagaac gcgtcgctga agcccaaagg tttgcccgaa 1260
tggagttaag ggagcatgga gaaaagaaga aacgctcgcg agaggatgct ggagataatg 1320
atgacacaga ggtgtctatt ggtgtcagga acaagggtggc tggaggaaaa atgaagaagc 1380
ggaaaggccg ttaatcactt ttatgaagcg tcgagttctg ctgttctgta aaagagaatt 1440
ggagaatgaa acctgctcca acagagatca tgagactgaa attggtcaga attgtgtcca 1500
gaatgtgctc agctaattca gtattcttcc ccattctggg ttggagtta ctgcagagta 1560
attcttacag tgctgatgct aagactgtta ctgttcttcg actttgatc cttgctcatg 1620
acatgagtag ggtgtgctct tctgtcactt cacacagacc ttttgccctt ttagctgca 1680
agtcaaggac taggttgatg atgccatga cctgtaattg taaagaagct tggacatctg 1740
caaatgatat ttaaaccatc ttggcttctg ctttattcaa actaatgtga aacaataaat 1800
ttaaataatta tttttaaaaa aaa 1823

```

&lt;210&gt; 13

&lt;211&gt; 869

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 13

```

cagcattgca gcagctccac catggcctgg gtcctctgc tcctcaccct cctcagtctc 60
ctcacagggt ccctctccca gcctatcttg actcagccac cttctgcac agcctccctg 120
ggagcctcgg tcacactcac gtgcagtgtg agcagcgact acaagaatct tgaagtggac 180
tggtttcagc agagaccagg gaagggcccc cgttttgtca tgcgagtggg cactgggtggc 240
gttgtgggat tcagaggggc tgacatccct gatcgctttt cagtctcggg ctcaggcctg 300
aatcggtttc tgaccatcag gaacatcgaa gaaggagatg agagtgacta ccactgtggg 360
acggaccttg gcagtgggac cagcttcgtg tcttgggtgt tcggcgaggg gaccaagttg 420
accgtcctaa gtcagcccaa ggctgcccc tcggtcactc tgttcccgcc ctctctgag 480
gagcttcaag ccaacaaggc cacactgggtg tgtctcataa gtgacttcta cccgggagcc 540
gtgacagtgg cctggaaggc agatagcagc cccgtcaagg cgggagtggg gaccaccaca 600
ccctccaaac aaagcaacaa caagtaacgg gccagcagct atctgagcct gacgcctgag 660
cagtggaggt ccaacagaag ctacagctgc caggtcacgc atgaaggag caccgtggag 720
aagacagtgg cccctacaga atgttcatag gttctaaacc ctcaccccc ctacgggaga 780
ctagagctgc aggatccag gggaggggtg tctcctccca cccgcaaggc gtcaagccct 840
tctccctgca ctcaataacc gatcgaata 869

```

&lt;210&gt; 14

&lt;211&gt; 799

&lt;212&gt; DNA





&lt;213&gt; Homo sapiens

&lt;400&gt; 14

```

ccctgctcag ctcttggggc cgctaattgct ctgggtccct gtgcagagat tgtgatgacc 60
cagactccac tctccttgct tatcaccctt ggagagcagg cctccatgtc ctgcagggtct 120
agtcagagcc tcttgcatag tgatggatac acctatttgt attggtttct gcagaaagcc 180
aggccagtct ccacagctcc tgatctatga agtttccaac cggttctctg gagtgtcacc 240
attaggttca gtggcagcgg gtccgggaga gaattcacat tgagaatcag ccgggtggag 300
gctgacgatg ctggagttta ctactgcatg caaactacac agactccgaa cacttttggc 360
caggggacga ggctggagat caaacgaact gtggctgcac catctgtctt catcttccc 420
ccatctgatg agcagttgaa atctggaact gcctctgttg tgtgcctgct gaataacttc 480
tatcccagag aggccaaagt acagtggaag gtggataacg ccctccaatc gggtaactcc 540
caggagagtg tcacagagca ggacagcaag gacagcacct acagcctcag cagcaccctg 600
acgctgagca aagcagacta cgagaaacac aaactctacg cctgcgaagt caccatcag 660
ggcctgagct cgcccgtcac aaagagcttc aacaggggag agtggttagag ggagaagtgc 720
ccccacctgc tctcagttc cagcctgacc ccctcccatc ctttggcctc tgaccctttt 780
tccacagggg acctacccc                                     799

```

&lt;210&gt; 15

&lt;211&gt; 1731

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 15

```

tttttttttt ttggttggtca ttgaggatat ttattggggt ttcattgagt cagggagaag 60
ggctggatga cttgggatgg ggagagagac ccctcccctg ggatcctgca gctccaggct 120
cccgtgggtg gggtagagt tgggaaccta tgaacattct gtaggggcca ctgtcttctc 180
cacggtgctc ccttcatgct tgacctggca gctgtagctt ctgtgggact tccactgctc 240
gggcgtcagg ctcaggtagc tgctggccgc gtacttggtt ttgctctgtt tggagggttt 300
ggtggtctcc actccgcct tgacggggct gccatctgcc ttccaggcca ctgtcacagc 360
tcccgggtag aagtcactga tcagacacac tagtggtggc ttgttggtt ggagctcctc 420
agaggagggc gggaacagag tgacagtggg gttggccttg ggctgacctg tgtggacagg 480
gaagggggtg agagagggca gacagaatac cggggtgttg tggagccctc ctctctgtct 540
aaagtctctg ggagggttca cagtgtggcc atccggtcca cccgggggtc tctcctcttc 600
tttccccatc ctttccactc atgcctgtg gagagcagac agctctgtgc cttcctagga 660
gccctcccaa gtcacctttc acagggtgtc tggcccagcc cctctcctct acagcctcaa 720
tttctccata taccagggc aggctgtgtt cttcttctc tgtgttctt cttctctgat 780
ctctggagtc tgagtttaga atctggcctt gaccctgga tccctcattt ccatccctat 840
acccccctcc atcacccact ttattcttcc aggacctaga gcctcctccg tgaactgtgg 900
gtcccctgct catcctgtag gactgtcctg gcaggcagtg tgtggggaga cccaagcctg 960
ctttgaactg gagtttcta tccctcacag gcaccgggct ctgcccagc ccagcccact 1020
tggtctctct ggcaaggagt gggctccacc tagacaagcc tcagggtcc tctgaggctc 1080
tgagatgttt cctgtctcca caagcccagc cacagaacct ttcacctcag ctgttctctg 1140
ggctgctccc acagactatg agactcagca gccccaggc ccaccccagc agcctgtgat 1200
gaccccaaac ttcacctggt agccatggag ttctctgcac ccctcattca ctcccctctg 1260
gtcatttctc gagtctaaca tgccctttga ggaaggcagg aggccgatcc gtgaacagag 1320
agacactggg cccagagggt gacggggctc cagggacaga cacatctctg ccctaagaga 1380

```



ctgtctcctt tctggtgact gtccctgggag ggttggattc tggcacctca cccagtctca 1440  
 cccgtccctc tgtgtcccca tgtcctcatg acccagcaca ggccacagag ctgcagccta 1500  
 gaccagagc cctctctgtg cctccattg gtctcccctt ggggtgacct ctgtgtcacc 1560  
 aggccgtgtg gccctcccag gctgattggc atccagtcct cagcctagac cctcagctgt 1620  
 tcttggggct gctcccctag actatgagac tcagcagctc ccaggcccac cccagcagcc 1680  
 tctgtccttg accccaggag tcaactggga atgtccctaa ggacactgca g 1731

<210> 16

<211> 662

<212> DNA

<213> Homo sapiens

<400> 16

gccaaagagt tccaggccca caagaggact ggctatgagg aagagacctg gaatctgaag 60  
 gaatgtgttg ggcgttgtgc aaaccctaac gtaaatttcc tgacaaaggt agaaagccct 120  
 ggcattggtc agaggtgggg cctcctccta tgtcgacggg attctagatt cacaccatgg 180  
 cagaaaatct atttcaggaa cttcaggaaac attttcaagc tctgacggca acattaaacc 240  
 tcagaaatat ccttttctac ctttaacaaa tgctgtgatt ctttcggact ggtagattat 300  
 catggagtat ctttttgttg tctggtagta gtaggtaata gtttacttag gatttcccag 360  
 tattttacttc tgtgctttta tgtggcttcc tgatgtgtta attaccctc acctatagca 420  
 aaagctgtac ctccggccgg gtgcagtggc ccataaggaa attttctaac tatatgaact 480  
 gaagggccat ggcttctaca gaaaatatgt ttacatcaaa tacaatcttg ggaagaataa 540  
 aaaatagctc tcctattcct tacagggaag gctataaaca atattttatt gtactgtttt 600  
 tataaccaga gtaaaccctt ggattctgtc atggattgaa ttgtgtaacc gcaaaattta 662  
 ta

<210> 17

<211> 336

<212> DNA

<213> Homo sapiens

<220>

<221> unsure

<222> (268)

<400> 17

tttaaaaaaa atccatctta gtatcttgac ccccaccctt caccactca cagagaagcc 60  
 cacatgagga aacagggtat gtcttggaca tctctgtccc cctcagtgtc tggatatagt 120  
 actgacacac agcatgttct caagaaatgt ttgaatcaca gtacattgaa tcagtaacag 180  
 tctgactgac cccaggcag aaaatgcaga ggcatttttt ctctctattc cagatttcag 240  
 ctgtagcctc ttgtaattct catattgntt ttcaatcacc agaattgatt tccctcatcc 300  
 ctcttcccag ggtcatctcc agtgaactgt attaatt 336

<210> 18

<211> 3300



&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; unsure

&lt;222&gt; (892)

&lt;400&gt; 18

gagccccgagc cgcgccaccc cgccctggcca tggcttttgc aagtttccgc cgcacccctgg 60  
 ccttgtctac cttcgagaag agaaagtccc gcgaatatga gcacgtccgc cgcgacctgg 120  
 accccaacga ggtgtgggag tatcgtgggc gagctgggcg acggcgtctt tcggaatggt 180  
 ttacaaggcc aagaataagg agacgggtgc tttggctgcg gcaatagtca ttgaaaccaa 240  
 gagtgaggag gagctggagg actacatcgt ggagattgag atcctggcca cctgcgacca 300  
 cccctacatt gtgaagctcc tgggagccta ctatcacgac gggaagctgt ggatcatgat 360  
 tgagttctgt ccagggggag ccgtggacgc catcatgctg gagctggaca gaggcctcac 420  
 ggagccccag atacagggtg tttgccgcca gatgctagaa gccctcaact tcctgcacag 480  
 caagaggatc atccaccgag atctgaaagc tggcaacgtg ctgatgaccc tcgagggaga 540  
 catcaggctg gctgactttg gtgtgtctgc caagaatctg aagactctac agaaacgaga 600  
 ttcccttcac ggcacgcctt actggatggc ccccgaggtg gtcattgtgtg agaccatgaa 660  
 agacacgccc tacgactaca aagccgacat ctggtccctg ggcattcacgc tgattgagat 720  
 ggcccagatc gagccgccac accacgagct caaccccatg cgggtccctg taaagatcgc 780  
 caagtcagac cctcccacgc tgctcacgcc ctccaagtgg tctgtagagt tccgtgactt 840  
 cctgaagata gccctggata agaaccaga aacccgaccc agtgccgcgc antgctggag 900  
 catcccttcg tcagcagcat caccagtaac aaggctctgc gggagctggt ggctgaggcc 960  
 aaggccgagg tgatggaaga gatcgaagac ggccgggatg agggggaaga ggaggacgcc 1020  
 gtggatgctg cctccaccct ggagaaccat actcagaact cctctgaggt ggtccgcca 1080  
 agcctcaatg ctgacaagcc tctcgaggag tcaccttcca ccccgctggc acccagccag 1140  
 tctcaggaca gtgtgaatga gccctgcagc cagccctctg gggacagatc cctccaaacc 1200  
 accagtcccc cagtcgtggc cccctggaaat gagaacggcc tggcagtgcc tgtgcccctg 1260  
 cggaagtccc gaccctgtgc aatggatgcc agaattcagg tagcccagga gaagcaagtt 1320  
 gctgagcagg gtggggacct cagcccagca gccaacagat ctcaaaaggc cagccagagc 1380  
 cggcccaaca gcagcgcctt ggagaccttg ggtggggaga agctggccaa tggcagcctg 1440  
 gagccacctg cccaggcagc tccagggcct tccaagaggg actcggactg cagcagcctc 1500  
 tgcacctctg agagcatgga ctatggtacc aatctctcca ctgacctgtc gctgaacaaa 1560  
 gagatgggct ctctgtccat caaggacctg aaactgtaca aaaaaacctc aagcggacac 1620  
 gcaaatttgt ggtggatggt gtggaggtga gcatcaccac ctccaagatc atcagcgaag 1680  
 atgagaagaa ggatgaggag atgagatttc tcaggcgcca ggaactccga gagcttcggc 1740  
 tgctccagaa agaagagcat cggaaccaga cccagctgag taacaagcat gagctgcagc 1800  
 tggagcaaat gcataaacgt tttgaacagg aaatcaacgc caagaagaag ttctttgaca 1860  
 cggaattaga gaacctggag cgtcagcaaa agcagcaagt ggagaagatg gagcaagacc 1920  
 atgccgtgcg ccgccgggag gaggccaggc ggatccgcct ggagcaggat cgggactaca 1980  
 ccaggttcca agagcagctc aaactgatga agaaagaggt gaagaacgag gtggagaagc 2040  
 tcccccgaca gcagcgggaag gaaagcatga agcagaagat ggaggagcac acgcagaaaa 2100  
 agcagcttct tgaccgggac tttgtagcca agcagaagga ggacctggag ctggccatga 2160  
 agaggctcac caccgacaac aggcggggaga tctgtgacaa ggagcgcgag tgcctcatga 2220  
 agaagcagga gtccttcga gaccgggaag cagccctgtg ggagatggaa gagcaccagc 2280  
 tgcaggagag gcaccagctg gtgaagcagc agctcaaaga ccagtacttc ctccagcggc 2340  
 acgagctgct gcgcaagcat gagaaggagc gggagcagat gcagcgctac aaccagcgca 2400



```

tgatagagca gctgaaggtg cggcagcaac aggaaaaggc ggggctgccc aagatccaga 2460
ggagtgaagg caagacgcgc atggccatgt acaagaagag cccctgttaa actggagttg 2520
ctggtgggca gccctgcttc ctgcatggaa ctggagctgt atggagttga cgacaagttc 2580
tacagcaagc tggatcaaga ggatgcgctc ctgggctcct accctgtaga tgacggctgc 2640
cgcatccacg tcattgacca cagtggcgcc cgccttggtg agtatgagga cgtgtccccg 2700
gtggagaagt acacgatctc acaagaagcc tacgaccaga ggcaagacac ggtccgctct 2760
ttcctgaagc gcagcaagct cggccggtac aacgaggagg agcgggctca gcaggaggcc 2820
gaggccgccc agcgcctggc cgaggagaag gccaggcca gctccatccc cgtgggcagc 2880
cgctgtgagg tgcgggcggc gggacaatcc cctcgccggg gcaccgcat gtatgtaggt 2940
ctcacagatt tcaagcctgg ctactggatt ggtgtccgct atgatgagcc actggggaaa 3000
aatgatggca gtgtgaatgg gaaacgctac ttogaatgcc aggccaagta tggcgcttt 3060
gtcaagccag cagtctgtgac ggtggggggac ttcccgagg aggactacgg gttggacgag 3120
atatgacacc taaggaattc ccctgcttca gctcctaget cagccactga ctgcccctcc 3180
tgtgtgtgcc catggccctt ttctcctgac cccattttaa ttttattcat ttttccctt 3240
gccattgatt tttgagactc atgcattaaa ttcactagaa acccagaaaa aaaaaaaaaa 3300

```

<210> 19

<211> 349

<212> DNA

<213> Homo sapiens

<220>

<221> unsure

<222> (6)

<220>

<221> unsure

<222> (14)

<220>

<221> unsure

<222> (25)

<220>

<221> unsure

<222> (28)

<220>

<221> unsure

<222> (30)

<220>

<221> unsure

<222> (32)

<220>

<221> unsure





<222> (34)

<220>

<221> unsure

<222> (54)

<220>

<221> unsure

<222> (60)

<220>

<221> unsure

<222> (66) .. (67)

<220>

<221> unsure

<222> (83)

<220>

<221> unsure

<222> (85)

<220>

<221> unsure

<222> (92)

<220>

<221> unsure

<222> (95)

<220>

<221> unsure

<222> (103) .. (104)

<220>

<221> unsure

<222> (145)

<220>

<221> unsure

<222> (152)

<220>

<221> unsure

<222> (160)

<220>

<221> unsure



<222> (179)

<220>

<221> unsure

<222> (197)

<220>

<221> unsure

<222> (222)

<220>

<221> unsure

<222> (238)

<220>

<221> unsure

<222> (246)

<220>

<221> unsure

<222> (249)

<220>

<221> unsure

<222> (258)

<220>

<221> unsure

<222> (266)

<220>

<221> unsure

<222> (273) .. (274)

<220>

<221> unsure

<222> (276)

<220>

<221> unsure

<222> (284)

<220>

<221> unsure

<222> (286)

<220>

<221> unsure



<222> (290)

<220>

<221> unsure

<222> (298)

<220>

<221> unsure

<222> (324)

<220>

<221> unsure

<222> (328)

<220>

<221> unsure

<222> (335)

<220>

<221> unsure

<222> (347)

<400> 19

ttaaanattt aaanatatta aggtntcntn tncngctcat cttcacagga aaanaattan 60  
 attccnnaca aacacctttc aantnttacc cnggnaagct gtnngaactg gaccttgaag 120  
 aaggaaatac tttctgactc attgnaattt gntatagtca nagtaagtaa ggcacaaant 180  
 ttgaagataa acttggnaca gtatgtagtg ccttttggtt tnaaatgcaa ctttggngaa 240  
 ttactngana ttttttcttc ataganttac tannanaagt tttnanaatn ttactgatna 300  
 ccactgtttt tgattttgct attnttttca aattnactat atatgangt 349

<210> 20

<211> 4665

<212> DNA

<213> Homo sapiens

<400> 20

agcggggagg gccccgagcg gcgcagatag ggaggttggg gctgtgcccc gcggcgcggc 60  
 gcctgccact gcgcaggcgc ctcaggaaga gctcggcacg gcccctcttc ctccagggtcc 120  
 cccttccccg caacttccca cgagtgccag gtgccgagcg cgccgagttc cgcgcatagg 180  
 aaagaagcga ccgcggcggc tggaaccctg attgctgtcc ttcaacgtgt tcattatgaa 240  
 gttattagta atacttttgt tttctggact tataactggg tttagaagtg actcttcctc 300  
 tagtttgcca cctaagttac tactagtatc ctttgatggc ttcagagctg attatctgaa 360  
 gaactatgaa tttcctcatc tccagaattt tatcaaagaa ggtgttttgg tagagcatgt 420  
 taataatggt tttatcacia aaacatttcc aaaccactac agtattgtga caggcttgta 480  
 tgaagaaagc catggcattg tggctaattc catgtatgat gcagtcacaa agaaacactt 540  
 ttctgactct aatgacaagg atcctttttg gtggaatgag gcagtcacaa tttgggtgac 600  
 caatcagctt caggaaaaca gatcaagtgc tgctgctatg tggcctggta ctgatgtacc 660



cattcacgat accatctctt cctatTTTTat gaattacaac tcctcagtgt catttgagga 720  
aagactaaat aatattacta tgtggctaaa caattcgaac ccaccagtca cctttgcaac 780  
actatatttg gaagaaccag atgcaagtgg ccacaaatag ggacctgaag ataaagaaaa 840  
catgagcaga gtgttgaaaa aaatagatga tcttatcggt gacttagtcc aaagactcaa 900  
gatgttaggg ctatgggaaa atcttaaatgt gatcattaca agtgcacatg ggatgaccca 960  
gtgttctcag gacagactga taaacctgga ttcttgcatc gatcattcat actacactct 1020  
tatagatttg agcccagttg ctgcaatact tcccaaaaata aatagaacag aggtttataa 1080  
caaactgaaa aactgtagcc ctcatatgaa tgtttatctc aaagaagaca ttcttaacag 1140  
atTTTTattac caacataatg atcgaattca gccattatt ttggttgccg atgaaggctg 1200  
gacaattgtg ctaaataaat catcacaaaa attaggtgac catggttatg ataattcttt 1260  
gcctagtatg catccatttc tagctgcca cggacctgca ttccacaaag gctacaagca 1320  
tagcacaaat aacattgtgg atatttatcc aatgatgtgc cacatcctgg gattaaaacc 1380  
acatcccaat aatgggacct ttggtcatatc taagtgttg ttagttgacc agtgggtgcat 1440  
taatctccca gaagccatcg cgattgttat cgggtcactc ttggtgttaa ccatgctaac 1500  
atgcctcata ataatacatg agaataagact ttctgtacct cgtccatttt ctgcacttca 1560  
gctacaagaa gatgatgatg atcctttaat tgggtgacat gtgctagggc ttatacaaaag 1620  
tgtctttgat taatcacaaa actaagaata catccaaaga atagtgttgt aactatgaaa 1680  
aagaatactt tgaaagacaa agaacttaga ctaagcatgt taaaattatt actttgtttt 1740  
ccttggtgtt tgtttcggtg catttgctaa taagataacg ctgaccatag taaaattgtt 1800  
agtaaatcat taggtaacat cttgtggtag gaaatcatta ggtaacatca atcctaacta 1860  
gaaatactaa aaatggcttt tgagaaaaat acttctctcg cttgtatttt gcgatgaaga 1920  
tgtgatacat ctttaaatga aaatatacca aaatttagta ggcatgtttt tctaataaat 1980  
ttatatattt gtaaagaaaa caacagaaat ctttatgcaa tttgtgaatt ttgtatatta 2040  
gggaggaaaa gcttctata tttttatatt tacctttaat tagtttgtat ctcaagtacc 2100  
ctcttgaggt aggaaatgct ctgtgatggt aaataaaatt ggagcagaca gaaaagatat 2160  
agcaaatgaa gaaatatattt aaggaaacct atttgaaaaa aaaagcaaag accatttgat 2220  
aaaagcctga gttgtcacca ttatgtctta agctgttagt cttaaagatt attgttaaaa 2280  
aattcagaag aaaagagaga caagtgtctt tctctctatc tatgcttaat gcctttatgt 2340  
aagttactta gttgtttgct tgtgctgtg caagtgtgtt tgtgtgtggt tgtgtggaca 2400  
ttatgtgatt tactatataa ggaggtcaga gatggactgt ggccaggctt ccacattcct 2460  
gaagcacaca gatctcagga aaggttattt ttgcacttca tatttgttta ctttctccta 2520  
actcacaagt taaaatcata acttaatttc attaaacttt atcatttaac tctctcatgt 2580  
ttgttgtaac ccgaggtatc caaatgctgc agaaaaattt atgacccaaa tacaaatctc 2640  
aatatgactg ggacagaatg aggaatggag atttttgtat ttatctttgg gactttatgc 2700  
cttacttttt aggtatata atagttaaga aattttaaac aaaatttagt atcttttggg 2760  
ctttcacacc attcatatgt taagtggcag aatagcctta gtgctacctc cacttttttc 2820  
tccagtattt gcatcacaga aataatccct ctgtttaaca tgtttgttca gagccaaggg 2880  
tttattgtga agaactgtca tcctgccttt gctagctggg accttctagt aatcaaaatt 2940  
aatatgaaga aactaggttg tgacagacta gattatattt agtaggggaa aaattgggct 3000  
caagaaccat tcatcagtac gtgagacaag cagttaatag tatgatcttt aaagttttga 3060  
caatataaaa taaacttggt aactgtttta caaatataaa agtataataa atatgcagcc 3120  
cagttaaata ttgattatct gtgatggtaa agaacaacag tggtgccagt catcaaact 3180  
acagtgcgtc ctattgagtc actgctaatt tcttgagcct ggtatttgcg gcctattgta 3240  
tttgtggttg ttgagaggca ttttcaaacc ctgtataaat aatccatgct gttggctata 3300  
agttaactgt attaagaaca gtaaaataaa taaaaaccaa tagtactaat tttgctttta 3360  
aaaaatttct aatttttttc acataaaaaca attatcctaa aggttaatat ttgatcgaaa 3420  
cagaataata gaaaaattct tctttaattt ccattaaaaa gcaaatagca ttgacacatt 3480  
taaagctttt catttaaaagt agtggatgtt tttgaagtat ctaaaatagt agcagaatat 3540





```

tttatacttg gtccttgcaa tgggtgtgagt tttaatgatt gcattatcgt gattgggtgg 3600
tatgagtttc agaaatctat acttggcatc caactcatga gtggatttta tataggatgg 3660
aacaggaagg tatgtcctgt cagtatctta accctttcaa caagacattt acctatttgt 3720
ctttccttac gttctcaaaa tattaactcg aattgtaaat taagcaaaaa tgtaaaaaagt 3780
atatgttgat gggacaagaa gaatagtatt tatttaataa aacatatatt atattgaact 3840
atgtgttaat tcatttgtat cttttaaaaa attatcactg ttaaagccat tgactccttt 3900
agtacactga gaaaaatctt atagtaaaac tagcctttca cattaagggt ttgggtgtgta 3960
ttttgttaaa taactaacat gctgctctat tttctgggtg tagaaagtat ttggctctag 4020
gaaacattta cttgtttgtg aaaacaatac cccaaggtaa taggaaaagt ttgagttaag 4080
tgtttttaat tcagtcagtg aattcagaat aagtacattc atgtataaca tagggacagt 4140
tctgctgctg ttatttataat gcaattcttc tggtaaatag caatagaata aaacatatatt 4200
caatgtttgt gtatagggtt tatattatta ttccactagg aatggcataa gaatttatag 4260
ataaattctt gtaacattaa aggattaaaa tgtttttaca ttgtttttgg gtgtctcctt 4320
cttgtgccca tatctgataa gctttatgga ttattgcatt taattccttt tatttgagg 4380
gttttacttc cttgttaaca tataaagtta taaatgaagg acaaggagga gatggaaaat 4440
gtgtatttat tgtaattct taaaatagtg tgtaataaaa ataacatcag tgtgctttaa 4500
agaaatgtgt atgtagtgcc ttaatttaaa ttaaaatatt tttgactgtt acttgagttc 4560
agaattaatg actttgttca tgatttttaa aatgtgtgtg aataaaatct accaaaaaat 4620
tcttactgta attattaaat ataaagttca gtgtcaaaaa aaaaa 4665

```

&lt;210&gt; 21

&lt;211&gt; 437

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 21

```

tagcaacagg cctggagggtg ctgcagtagt gggggaaaat ggaagggtgga ggggtggagt 60
tttgctgcag gacagctgag tggagggttg ggacaggtgc aaactggaga ggcctagaga 120
gctagagaag caagtaaggg ccagggccag agtcggcttc aatggaaca cagccagtg 180
ccctaaggcc cctaactctt gctggctgtt tcttgacccc aagccagggt tgggagtcct 240
ctgggcatcc attttttcta aaggaactgg acagagtaca cacaggaaag gaagctgtca 300
ccctcttgcc atctggctcc aggggcctcc agtcagcat tcctccttct tcccttgatt 360
gggtggggcc acatgatggg cagccaggct ctgggctgtc ccactagagc agctgcaaac 420
acagccatgt ttcagt 437

```

&lt;210&gt; 22

&lt;211&gt; 355

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 22

```

caagctggct tctgttnaga tgagctncnn ggagatgcta ctgcatggca caggaagaac 60
gtgtgaccac taggattatt tccagcataa ntggctttgc atggntgaag ttntagcaat 120
gaatttctat aagccctttt aaaattggaa ttcataaaca agtctctgtg ctctcacnt 180
gtggcanttn tttctgctct tttgtgtgt tnatgtgtt ctactgcta cctagctagc 240
atcntgggtg catggaagtg gaccagatat ttccacaccc attatattct agatgctgtg 300

```



ttaagatnca ngacaacnan ntngnnncnn gatgtaaaat tttntagncn gnagg 355

<210> 23

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic

<400> 23

ccagagccca aatcttgtga c

21

<210> 24

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic

<400> 24

gcggctttgt cttggcatta

20

<210> 25

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic

<400> 25

attgccatcc cagtgcagct g

21

<210> 26

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic

<400> 26



21

ttgggagatg tgggtgatga g

&lt;210&gt; 27

&lt;211&gt; 22

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Description of Artificial Sequence: Synthetic

&lt;400&gt; 27

22

cctgccctgg tatgtttttc tt

&lt;210&gt; 28

&lt;211&gt; 21

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Description of Artificial Sequence: Synthetic

&lt;400&gt; 28

21

cagcccacaa atgccttcta c

&lt;210&gt; 29

&lt;211&gt; 24

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Description of Artificial Sequence: Synthetic

&lt;400&gt; 29

24

ccactaggat tatttccagc ataa

&lt;210&gt; 30

&lt;211&gt; 24

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Description of Artificial Sequence: Synthetic

&lt;400&gt; 30



24

ggtgtgaaaa tatctggtcc actt

<210> 31

<211> 18

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic

<400> 31

18

agccattgcc atcccagt

<210> 32

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic

<400> 32

20

atgttcttca cgctcttcgc

<210> 33

<211> 24

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic

<400> 33

24

aggaagtgcg ggaagaggct ggct

<210> 34

<211> 19

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic

<400> 34





19

aagggagcac cgtggagaa

&lt;210&gt; 35

&lt;211&gt; 19

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Description of Artificial Sequence: Synthetic

&lt;400&gt; 35

19

agggctggat gacttggga

&lt;210&gt; 36

&lt;211&gt; 24

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Description of Artificial Sequence: Synthetic

&lt;400&gt; 36

24

ttcccaactc taaccccacc cacg

&lt;210&gt; 37

&lt;211&gt; 7444

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 37

gtctcctctg gatcttaact actgagcgca atgctgagcc atggagccgg gttggccttg 60  
 tggatcacac tgagcctgct gcagactgga ctggcggagc cagagagatg taacttcacc 120  
 ctggcggagt ccaaggcctc cagccattct gtgtctatcc agtggagaat tttgggctca 180  
 ccctgtaact ttagcctcat ctatagcagt gacaccctgg gggccgcgtt gtgccctacc 240  
 tttcggatag acaacaccac atacggatgt aaccttcaag atttacaagc aggaaccatc 300  
 tataacttca ggattatttc tctggatgaa gagagaacag tggctcttgca aacagatcct 360  
 ttacctcctg ctaggttttg agtcagtaaa gagaagacga cttcaaccag cttgcatgtt 420  
 tggtggaactc cttcttccgg aaaagtcacc tcatatgagg tgcaattatt tgatgaaaat 480  
 aacccaaaaga tacaggggggt tcaaattcaa gaaagtactt catggaatga atacactttt 540  
 ttcaatctca ctgctggtag taaatacaat attgccatca cagctgtttc tggaggaaaa 600  
 cgttcttttt cagtttatac caatggatca acagtgccat ctccagtgaag agatattggg 660  
 atttccacaa aagccaattc tctcctgatt tcctgggtccc atgggtcttg gaatgtggaa 720  
 cgataaccggc tgatgctaata ggataaagggt atcctagtgc atggcggtgt tgtggacaaa 780  
 catgctactt cctatgcttt tcacgggctg tcccctggct acctctacaa cctcactgtt 840  
 atgactgagg ctgcagggtc gcaaaactac aggtggaaac tagtcaggac agcccccatg 900



gaagtctcaa atctgaaggt gacaaatgat ggcagtttga cctctctaaa agtcaaattgg 960  
 caaagacctc ctggaaatgt ggattcttac aatatcacc cttgtctcaca agggaccatc 1020  
 aaggaatcca gagtattagc accttggtt actgaaactc actttaaga gttagtcccc 1080  
 ggtcgacttt atcaagttac tgtcagctgt gtctctggtg aactgtctgc tcagaagatg 1140  
 gcagtgggca gaacatttcc agacaaagt gcaaacctgg aggcaaaca taatggcagg 1200  
 atgaggtctc ttgtagttag ctggctgccc cctgctggag actgggagca gtatcggtac 1260  
 ctactcttca atgattctgt ggtgctgctc aacatcactg tgggaaagga agaaacacag 1320  
 tatgtcatgg atgacacggg gctcgtaccg ggaagacagt atgaggtgga agtcattgtt 1380  
 gagagtggaa atttgaagaa ttctgagcgt tgccaaggca ggacagtccc cctggctgtc 1440  
 ctccagcttc gtgtcaaaca tgccaatgaa acctcactga gtatcatgtg gcagaccctt 1500  
 gtagcagaat gggagaaata catcatttcc ctactgtgaca gagacctctt actgatccac 1560  
 aagtcactct ccaaagatgc caaagaattc acttttactg acctggtgcc tggacgaaaa 1620  
 tacatggcta cagtcaccag tattagtggg gacttaaaaa attcctcttc agtaaaagga 1680  
 agaacagtgc ctgcccagt gactgacttg catgtggcca accaaggaat gaccagttagt 1740  
 ctgtttacta actggaccca ggcacaagga gacgtagaat tttaccaagt cttactgatc 1800  
 catgaaaatg tggtcattaa aaatgaaagc atctccagt agaccagcag atacagcttc 1860  
 cactctctca agtccggcag cctgtactcc gtggtggtaa caacagttag tggagggatc 1920  
 tcttcccgac aagtggttgt ggagggaga acagtcctt ccagtgtgag tggagtaacg 1980  
 gtgaacaatt ccggtcgtaa tgactacctc agcgtttcct ggctcgtggc gcccgagat 2040  
 gtggataact atgaggtaac attgtctcat gacggcaagg tggttcagtc ccttgtcatt 2100  
 gccagtctg tcagagaatg ttccttcagc tccctcacc caggccgcct ctacaccgtg 2160  
 accataacta caaggagtgg caagtatgaa aatcactcct tcagccaaga gcgacagtg 2220  
 cctgacaaaag tccaggaggt cagtgttagc aactcagcca ggagtgacta ttttaagggt 2280  
 tcttgggtgt atgccactgg agactttgat cactatgaag tcaccattaa aaacaaaaac 2340  
 aacttcattc aaactaaaag cattcccaag tcagaaaacg aatgtgtatt tgttcagcta 2400  
 gtccctggac ggttgtacag tgtcactgtt actacaaaaa gtggacaata tgaagccaat 2460  
 gaacaaggga atgggagaac aattccagag cctgttaagg atctaacatt gcgcaacagg 2520  
 agcactgagg acttgcatgt gacttggta ggagctaag gggatgtcga ccaatatgag 2580  
 atccagctgc tcttcaatga catgaaagta tttcctcctt ttcaccttgt aaataccgca 2640  
 accgagtatc gatttacttc cctaaccacca ggccgccaat acaaaattct tgtcttgacg 2700  
 attagcgggg atgtacagca gtcagccttc attgagggtt tcacagttcc tagtgctgtc 2760  
 aaaaatatcc acatttctcc caatggagca acagatagcc tgacggtgaa ctggactcct 2820  
 ggtgggggag acgttgattc ctacacgggtg tgggcattca ggcacagtca aaagggtgac 2880  
 tctcagacta ttcccaagca cgtctttgag cacacgttcc acagactgga ggccggggag 2940  
 cagtaccaga tcatgattgc ctacgtcagc ggggtccctga agaactcagat aaatgtgggt 3000  
 gggcgggacag ttccagcatc tgtccaagga gtaattgcag acaatgcata cagcagttat 3060  
 tctttaatag taagtggca aaaagctgct ggtgtggcag aaagatatga tatcctgctt 3120  
 ctaactgaaa atggaatcct tctgcgcaac acatcagagc cagccaccac taagcaaac 3180  
 aaatttgaag atctaacc accgcaagaaa tacaagatac agatcctaac tgtcagtgg 3240  
 ggcctcttta gcaaggaagc ccagactgaa ggccgaacag tccagcagc tgtcaccgac 3300  
 ctgaggatca cagagaactc caccaggcac ctgtccttcc gctggaccgc ctgagagggg 3360  
 gagctcagct ggtacaacat ctttttgtac aaccagatg ggaatctcca ggagagagct 3420  
 caagttgacc cactagtcca gagcttctct ttccagaact tgctacaagg cagaatgtac 3480  
 aagatggtga ttgtaactca cagtggggag ctgtctaatt agtctttcat atttggtaga 3540  
 acagtcccag cctctgtgag tcatctcagg ggtccaatc ggaacacgac agacagcctt 3600  
 tggttcaact ggagtccagc ctctggggac tttgactttt atgagctgat tctctataat 3660  
 cccaatggca caaagaagga aaactggaaa gacaaggacc tgacggagtg gcggtttcaa 3720  
 ggccttgctc ctggaaggaa gtacgtgctg tgggtggtaa ctcacagtgg ggatctcagc 3780



aataaagtca cagcggagag cagaacagct ccaagtcctc ccagtcttat gtcatttgct 3840  
gacattgcaa acacatcctt ggccatcacg tggaaagggc cccagactg gacagactac 3900  
aacgactttg agctgcagtg gttgcccaga gatgcaacta ctgtcttcaa cccctacaac 3960  
aacagaaaaat cagaaggacg cattgtgtat ggtcttcgtc cagggagatc ctatcaattc 4020  
aacgtcaaga ctgtcagtg tgattcctgg aaaacttaca gcaaaccaat ttttggatct 4080  
gtgaggacaa agcctgacaa gatacaaac ctgcattgcc ggcctcagaa ctccacggcc 4140  
attgcctgtt cttggatccc tcctgattct gactttgatg gttatagtat tgaatgccgg 4200  
aaaatggaca cccaagaagt tgagttttcc agaaagctgg agaaagaaaa atctctgctc 4260  
aacatcatga tgctagtgcc ccataagagg tacctgggtg ccatcaaagt gcagtcggcc 4320  
ggcatgacca gcgaggtggg tgaagacagc actatcacia tgatagaccg cccccctcct 4380  
ccacccccac acattcgtgt gaatgaaaag gatgtgctaa ttagcaagtc ttccatcaac 4440  
tttactgtca actgcagctg gttcagcgac accaatggag ctgtgaaata cttcacagt 4500  
gtgggtgagag aggtgatgg cagtgatgag ctgaagccag agcagcagca ccctctccct 4560  
tcctacctgg agtacaggca caatgcctcc attcgggtgt atcagactaa ttattttgcc 4620  
agcaaatgtg ccgaaaatcc taacagcaac tccaagagtt ttaacattaa gcttgagca 4680  
gagatggaga gcctaggtgg aaaatgcat cccactcagc aaaaattctg tgatggacca 4740  
ctgaagccac aactgccta cagaatcagc attcagctt ttacacagct ctttgatgag 4800  
gacctgaagg aattcacaaa gccactctat tcagacacat tttttcttt acccatcact 4860  
actgaatcag agcccttgtt tggagctatt gaaggtgtga gtgctggtct gtttttaatt 4920  
ggcatgctag tggctgttgt tgccttattg atctgcagac agaaagttag ccatggctga 4980  
gaaagaccct ctgcccgtct gagcattcgt agggatcgac cattatctgt ccacttaaac 5040  
ctgggccaga aaggttaaccg gaaaacttct tgtccaataa aaataaatca gtttgaagg 5100  
catttcatga agctacaggc tgactccaac taccttctat ccaaggaata cgaggagtt 5160  
aaagacgtgg gccgaaacca gtcattgtgac attgcactct tgccggagaa tagagggaaa 5220  
aatcgataca acaatatatt gccctatgat gccacgcgag tgaagctctc caatgtagat 5280  
gatgatcctt gctctgacta catcaatgcc agctacatcc ctggcaacaa cttcagaaga 5340  
gaatacattg tcaactcagg accgcttctt ggcaccaagg atgacttctg gaaaatggtg 5400  
tgggaacaaa acgttcacaa catcgtcatg gtgaccaggt gtgttgagaa gggccgagta 5460  
aagtgtgacc attactggcc agcggaccag gattccctct actatgggga cctcatcctg 5520  
cagatgctct cagagtcctt cctgcctgag tggaccatcc gggagttaa gatatgcgg 5580  
gaggaacagc ttgatgcaca cagactcatc cgccacttct actatacggg gtggccagac 5640  
catggagtcc cagaaaccac ccagtctctg atccagtttg tgagaactgt cagggactac 5700  
atcaacagaa gccggggtgc tgggcccact gtggtgact gcagtgtggt tgtgggtagg 5760  
actggaacct ttattgcatt ggaccgaatc ctccagcagt tagactcaa agactctgtg 5820  
gacatttatg gagcagtga cgacctaa cttcacaggg ttcacatggt ccagactgag 5880  
tgtcagtatg tctacctaca tcagtgtgta agagatgtcc tcagagcaag aaagctacgg 5940  
agtgaacaag aaaaccctt gtttccaatc tatgaaaatg tgaatccaga gtatcacaga 6000  
gatccagtct attcaaggca ttgagaatgt acctgaagag ctccctggata aaaattattc 6060  
actgtgtgat ttgtttttaa aaacttgctt catgccctac agagggtgcca gctatttctg 6120  
ttgatactat gtataattta ttaactctga gaatgtttaa aattttatat aatttaaagg 6180  
taacagatat tattgtacat agttgtattt tgtagtttct tctgtaaata tgtatttttc 6240  
ataatgttta atattaagct ttatataata ctatttttcc aactaaagt gttcatgact 6300  
tgttctacat aaaactaatt caacctgtat gacaggacta ctggtaaaat gcatatggag 6360  
gtggtggcag agacaatcct tcaggccatg ttttctacct gttttgatat tcaactggaca 6420  
ggaaaaaggg cagggtctaga gagagcaaat actatggcta gatttgcctg attgctgtcc 6480  
catgaatctc gagagccaac agacatgtcc taacttgcta ttaggacaaa tgtgacagtc 6540  
aaaaaaagga ttagagggag ggagaaaaaa gaattaagca gtaccaaagc tgaactagat 6600  
gcttgtgtct gaactagttg ctctctctct cctttctcct tccagggtt caagccaaag 6660



```

tggtcagctc agggatcatg taacttgag tgcaagccca ggatggtagg atgcagggtt 6720
gaggggtctg atagagaatg attccaaaca gaagtgatga attccttttg ttaataagat 6780
gccagctata ccagactgg aacataaca tgcaaagcac tatctacagt gattagagat 6840
cctttcattg cattcatggt gtggagtgtg aacatccaca ccataactgt aatgtattta 6900
tacacactag tttctgtctc attttcagt gtctccattc ctagaaaagt cacaattatc 6960
cattcctact tgatttccca ttaaaagaat attatggtag cagattgtgc ccctcattaa 7020
aaggcttaat gccaacattt tcatagaaat gactacaaac atcatatata gttaaattta 7080
aaacaatagc aaaaacaaaa acagtgggtc ttagtaaaat tttcaaaact tcttttagta 7140
aatcaatgaa gtcaaaatgt caagtaatca cccaaagttg catttaataa caaaaggcac 7200
tacatactgt accaagttta tcttcaatat ttgtgcctta cttactttga ctataacaaa 7260
ttccaatgag tcagaaagta tttccttcat caaggtccag ttccgacagc attcctggga 7320
aaaatttgaa aggagtgtgt acggaatctt catagatacc tgagaagatg agctggagat 7380
gtttgccttt ttcacactac aaatttttct gtaataaaact tgggaattag aggtcaaaaa 7440
aaaa

```

&lt;210&gt; 38

&lt;211&gt; 2475

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 38

```

atcacctgca tctcagagga cagacctgt gaagtcagag ctgctacaca ttgaatctca 60
agtcagactt ctgagattcg atgattcagg aagaaaggat tctgaggttt tgaagcaaaa 120
tgcagtgaac agcaaccaat ccaatgttgt aattgaagac tttgagtcct cacgatctc 180
ttcgtctttg cagcgtagcc cgagtcgggc agcgccggag gacctcagca gccatgtcga 240
agcccatag tgaagccggg actgccttca ttcagaccca gcagctgcac gcagccatgg 300
ctgacacatt cctggagcac atgtgccgcc tggacattga ttcaccaccc atcacagccc 360
ggaacactgg catcatctgt accattgggc ccagcttccc gatcagtggg gacgttgaag 420
gagatgataa gtctggaatg aatgtggctc gtctgaattc tctcatggac tcatgagtac 480
catgcggaga ccatcaagaa tgtgcgcaca gccacggaaa gctttgcttc tgaccccatc 540
ctctaccggc ccgttgctgt ggctctagac actaaaggac ctgagatccg aactgggctc 600
atcaagggca gcggcactgc agaggtggag ctgaagaagg gagccactct caaaatcacg 660
ctggataacg cctacatgga aaagtgtgac gagaacatcc tgtggctgga ctacaagaac 720
atctgcaagg tgggtggaag gggcagcaag atctacgtgg atgatgggct tatttctctc 780
caggtgaagc agaaagggtc cgacttctgt gtgacggagg tggaaaatgg tggctccttg 840
ggcagcaaga aggggtgtgaa ccttcctggg gctgctgtgg acttgctgc tgtgtcggag 900
aaggacatcc caggatctga aagtttggg gtcgagcagg atgttgatat ggtgtttggc 960
gtcattccat ccgcaaagg catctggatg tcccatggaa ngtttaggaa nggtcctggg 1020
gagagaaggg aaaagaaaca tccaagatta tccagcaaaa tcgagaatca tgaggggggt 1080
cggaggtttg atgaaatcct ggaggccagt gatgggatca tgggtggctc tggatgacta 1140
ggcattgaga ttcctgcaga gaaggtcttc ttgctcagaa gatgatgatt ggacgggtgca 1200
accgagact gggaagcctg tcatctgtgc tactccagat gctggagagc atcgatcaag 1260
aagccccgcc ccactcgggc tgaaggcagt gatgtggcca atgcagtcct ggatggagcc 1320
gactgcatca tgctgtctgg agaaacagcc aaaggggact atcctctgga ggctgtgcgc 1380
atgcagcacc tgattgccc tgaggcagag gctgccatct accacttgca attatttgag 1440
gaactccgcc gcctggcgcc cattaccagc gacccacag aagccaccgc cgtgggtgcc 1500
gtggaggcct ccttcaagt ctgcagtggg gccataatcg tctcaccaa gtctggcagg 1560

```





```

tctgtcacc aggtggccag ataccgcca cgtgccccca tcattgctgt gaccgcgaat 1620
ccccagacag ctgcgcaggc ccacctgtac cgtggcatct tccctgtgct gtgcaaggac 1680
ccagtccagg aggcctgggc tgaggacgtg gacctccggg tgaactttgc catgaatgtt 1740
ggcaaggccc gaggttctt caagaaggga gatgtggtca ttgtgctgac cggatggcgc 1800
cctggctccg gcttcaccaa caccatgcgt gttgttctctg tgccgtgatg gaccccagag 1860
cccctcctcc agcccctgtc ccaccccctt ccccagccc atccattagg ccagcaacgc 1920
ttgtagaact cactctgggc tgtaacgtgg cactggtagg ttgggacacc agggaagaag 1980
atcaacgcct cactgaaaca tggctgtgtt tgcagcctgc tctagtggga cagcccagag 2040
cctggctgcc catcatgtgg cccaccccaa tcaagggaag aaggaggaat gctggactgg 2100
aggcccctgg agccagatgg caagagggtg acagcttcct ttctgtgtg tactctgtcc 2160
agttccttta gaaaaaatgg atgccagag gactcccaac cctggcttgg ggtcaagaaa 2220
cagccagcaa gagttagggg ccttagggca ctgggctgtt gttccattga agccgactct 2280
ggccctggcc ctacttgct tctctagctc tctaggctc tccagtttgc acctgtcccc 2340
accctccact cagctgtcct gcagcaaaca ctccaccctc caccttccat tttccccac 2400
tactgcagca cctccaggcc tgttgctata gagcctacct gtatgtcaat aaacaacagc 2460
tgaagcacca aaaaa                                     2475

```

&lt;210&gt; 39

&lt;211&gt; 436

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 39

Glu Trp Ile Ser Tyr Ile Arg Gly Gly Gly Glu Arg Ile Tyr Tyr Ala

1

5

10

15

Asp Ser Val Arg Gly Arg Phe Thr Val Ser Arg Asp Asn Ala Lys Asn

20

25

30

Ser Leu Tyr Leu Gln Met Asn Ser Leu Arg Ala Glu Asp Thr Ala Val

35

40

45

Tyr Phe Cys Ala Arg Glu Pro Pro Ala Pro Asn Tyr Phe Asp Cys Trp

50

55

60

Ser Gln Gly Thr Leu Val Thr Val Ser Ser Ala Ser Thr Lys Gly Pro

65

70

75

80

Ser Val Phe Pro Leu Ala Pro Cys Ser Arg Ser Thr Ser Gly Gly Thr

85

90

95

Ala Ala Leu Gly Cys Leu Val Lys Asp Tyr Phe Pro Glu Pro Val Thr

100

105

110

Val Ser Trp Asn Ser Gly Ala Leu Thr Ser Gly Val His Thr Phe Pro

115

120

125



Ala Val Leu Gln Ser Ser Gly Leu Tyr Ser Leu Ser Ser Val Val Thr  
 130 135 140

Val Pro Ser Ser Ser Leu Gly Thr Gln Thr Tyr Thr Cys Asn Val Asn  
 145 150 155 160

His Lys Pro Ser Asn Thr Lys Val Asp Lys Arg Val Glu Leu Lys Thr  
 165 170 175

Pro Leu Gly Asp Thr Thr His Thr Cys Pro Arg Cys Pro Glu Pro Lys  
 180 185 190

Ser Cys Asp Thr Pro Pro Pro Cys Pro Arg Cys Pro Glu Pro Lys Ser  
 195 200 205

Cys Asp Thr Pro Pro Pro Cys Pro Arg Cys Pro Ala Pro Glu Leu Leu  
 210 215 220

Gly Gly Pro Ser Val Phe Leu Phe Pro Pro Lys Pro Lys Asp Thr Leu  
 225 230 235 240

Met Ile Ser Arg Thr Pro Glu Val Thr Cys Val Val Val Asp Val Ser  
 245 250 255

His Glu Asp Pro Glu Val Gln Phe Lys Trp Tyr Val Asp Gly Val Glu  
 260 265 270

Val His Asn Ala Lys Thr Lys Pro Arg Glu Glu Gln Phe Asn Ser Thr  
 275 280 285

Phe Arg Val Val Ser Val Leu Thr Val Leu His Gln Asp Trp Leu Asn  
 290 295 300

Gly Lys Glu Tyr Lys Cys Lys Val Ser Asn Lys Ala Leu Pro Ala Pro  
 305 310 315 320

Ile Glu Lys Thr Ile Ser Lys Thr Lys Gly Gln Pro Arg Glu Pro Gln  
 325 330 335

Val Tyr Thr Leu Pro Pro Ser Arg Glu Glu Met Thr Lys Asn Gln Val  
 340 345 350

Ser Leu Thr Cys Leu Val Lys Gly Phe Tyr Pro Ser Asp Ile Ala Val  
 355 360 365

Glu Trp Glu Ser Ser Gly Gln Pro Glu Asn Asn Tyr Asn Thr Thr Pro  
 370 375 380



Pro Met Leu Asp Ser Asp Gly Ser Phe Phe Leu Tyr Ser Lys Leu Thr  
 385 390 395 400

Val Asp Lys Ser Arg Trp Gln Gln Gly Asn Ile Phe Ser Cys Ser Val  
 405 410 415

Met His Glu Ala Leu His Asn Arg Phe Thr Gln Lys Ser Leu Ser Leu  
 420 425 430

Ser Pro Gly Lys  
 435

<210> 40

<211> 168

<212> PRT

<213> Homo sapiens

<400> 40

Pro Gln Leu Ala Cys Leu Phe Gln Val Lys Ser Gly Ser Pro Ala Val  
 1 5 10 15

Leu Ala Phe Ala Lys Glu Lys Ser Phe Gly Trp Pro Ser Phe Ile Thr  
 20 25 30

Tyr Thr Val Gly Val Ser Asp Pro Ala Ala Gly Ser Gln Gly Pro Leu  
 35 40 45

Ser Thr Thr Leu Thr Phe Ser Ser Pro Val Thr Asn Gln Ala Ile Ala  
 50 55 60

Ile Pro Val Thr Val Ala Phe Val Met Asp Arg Arg Gly Pro Gly Pro  
 65 70 75 80

Tyr Gly Ala Ser Leu Phe Gln His Phe Leu Asp Ser Tyr Gln Val Met  
 85 90 95

Phe Phe Thr Leu Phe Ala Leu Leu Ala Gly Thr Ala Val Met Ile Ile  
 100 105 110

Ala Tyr His Thr Val Cys Thr Pro Arg Asp Leu Ala Val Pro Ala Ala  
 115 120 125

Leu Thr Pro Arg Ala Ser Pro Gly His Ser Pro His Tyr Phe Ala Ala  
 130 135 140

Ser Ser Pro Thr Ser Pro Asn Ala Leu Pro Pro Ala Arg Lys Ala Ser  
 145 150 155 160



Pro Pro Ser Gly Leu Trp Ser Pro  
165

<210> 41

<211> 78

<212> PRT

<213> Homo sapiens

<400> 41

Val Ser Glu Gly Ala Thr Trp Ala Ile Gly Phe Pro Ala Ser Phe Pro  
1 5 10 15

Leu Phe Leu Ala Pro Ala Ala Glu Ala Gly Arg Pro Trp Arg Thr Ser  
20 25 30

Trp Gly Leu Thr Ala Ala Ser Pro Gly Ser Ser Trp Gly His Leu Ser  
35 40 45

Ser Lys Val Cys Thr Gln Glu Val Pro His His Ile Gln Pro His Gly  
50 55 60

Ser Pro Arg Ser Ala Arg Gln Gln Ile Arg Ala Pro Cys His  
65 70 75

<210> 42

<211> 1118

<212> PRT

<213> Homo sapiens

<400> 42

Met Ala Arg Ser Pro Gly Arg Ala Tyr Ala Leu Leu Leu Leu Ile  
1 5 10 15

Cys Phe Asn Val Gly Ser Gly Leu His Leu Gln Val Leu Ser Thr Arg  
20 25 30

Asn Glu Asn Lys Leu Leu Pro Lys His Pro His Leu Val Arg Gln Lys  
35 40 45

Arg Ala Trp Ile Thr Ala Pro Val Ala Leu Arg Glu Gly Glu Asp Leu  
50 55 60

Ser Lys Lys Asn Pro Ile Ala Lys Ile His Ser Asp Leu Ala Glu Glu  
65 70 75 80





Arg Gly Leu Lys Ile Thr Tyr Lys Tyr Thr Gly Lys Gly Ile Thr Glu  
 85 90 95

Pro Pro Phe Gly Ile Phe Val Phe Asn Lys Asp Thr Gly Glu Leu Asn  
 100 105 110

Val Thr Ser Ile Leu Asp Arg Glu Glu Thr Pro Phe Phe Leu Leu Thr  
 115 120 125

Gly Tyr Ala Leu Asp Ala Arg Gly Asn Asn Val Glu Lys Pro Leu Glu  
 130 135 140

Leu Arg Ile Lys Val Leu Asp Ile Asn Asp Asn Glu Pro Val Phe Thr  
 145 150 155 160

Gln Asp Val Phe Val Gly Ser Val Glu Glu Leu Ser Ala Ala His Thr  
 165 170 175

Leu Val Met Lys Ile Asn Ala Thr Asp Ala Asp Glu Pro Asn Thr Leu  
 180 185 190

Asn Ser Lys Ile Ser Tyr Arg Ile Val Ser Leu Glu Pro Ala Tyr Pro  
 195 200 205

Pro Val Phe Tyr Leu Asn Lys Asp Thr Gly Glu Ile Tyr Thr Thr Ser  
 210 215 220

Val Thr Leu Asp Arg Glu Glu His Ser Ser Tyr Thr Leu Thr Val Glu  
 225 230 235 240

Ala Arg Asp Gly Asn Gly Glu Val Thr Asp Lys Pro Val Lys Gln Ala  
 245 250 255

Gln Val Gln Ile Arg Ile Leu Asp Val Asn Asp Asn Ile Pro Val Val  
 260 265 270

Glu Asn Lys Val Leu Glu Gly Met Val Glu Glu Asn Gln Val Asn Val  
 275 280 285

Glu Val Thr Arg Ile Lys Val Phe Asp Ala Asp Glu Ile Gly Ser Asp  
 290 295 300

Asn Trp Leu Ala Asn Phe Thr Phe Ala Ser Gly Asn Glu Gly Gly Tyr  
 305 310 315 320

Phe His Ile Glu Thr Asp Ala Gln Thr Asn Glu Gly Ile Val Thr Leu  
 325 330 335



Ile Lys Glu Val Asp Tyr Glu Glu Met Lys Asn Leu Asp Phe Ser Val  
 340 345 350

Ile Val Ala Asn Lys Ala Ala Phe His Lys Ser Ile Arg Ser Lys Tyr  
 355 360 365

Lys Pro Thr Pro Ile Pro Ile Lys Val Lys Val Lys Asn Val Lys Glu  
 370 375 380

Gly Ile His Phe Lys Ser Ser Val Ile Ser Ile Tyr Val Ser Glu Ser  
 385 390 395 400

Met Asp Arg Ser Ser Lys Gly Gln Ile Ile Gly Asn Phe Gln Ala Phe  
 405 410 415

Asp Glu Asp Thr Gly Leu Pro Ala His Ala Arg Tyr Val Lys Leu Glu  
 420 425 430

Asp Arg Asp Asn Trp Ile Ser Val Asp Ser Val Thr Ser Glu Ile Lys  
 435 440 445

Leu Ala Lys Leu Pro Asp Phe Glu Ser Arg Tyr Val Gln Asn Gly Thr  
 450 455 460

Tyr Thr Val Lys Ile Val Ala Ile Ser Glu Asp Tyr Pro Arg Lys Thr  
 465 470 475 480

Ile Thr Gly Thr Val Leu Ile Asn Val Glu Asp Ile Asn Asp Asn Cys  
 485 490 495

Pro Thr Leu Ile Glu Pro Val Gln Thr Ile Cys His Asp Ala Glu Tyr  
 500 505 510

Val Asn Val Thr Ala Glu Asp Leu Asp Gly His Pro Asn Ser Gly Pro  
 515 520 525

Phe Ser Phe Ser Val Ile Asp Lys Pro Pro Gly Met Ala Glu Lys Trp  
 530 535 540

Lys Ile Ala Arg Gln Glu Ser Thr Ser Val Leu Leu Gln Gln Ser Glu  
 545 550 555 560

Lys Lys Leu Gly Arg Ser Glu Ile Gln Phe Leu Ile Ser Asp Asn Gln  
 565 570 575

Gly Phe Ser Cys Pro Glu Lys Gln Val Leu Thr Leu Thr Val Cys Glu  
 580 585 590



WO 02/08278

Cys Leu His Gly Ser Gly Cys Arg Glu Ala Gln His Asp Ser Tyr Val  
 595 600 605

Gly Leu Gly Pro Ala Ala Ile Ala Leu Met Ile Leu Ala Phe Leu Leu  
 610 615 620

Leu Leu Leu Val Pro Leu Leu Leu Leu Met Cys His Cys Gly Lys Gly  
 625 630 635 640

Ala Lys Gly Phe Thr Pro Ile Pro Gly Thr Ile Glu Met Leu His Pro  
 645 650 655

Trp Asn Asn Glu Gly Ala Pro Pro Glu Asp Lys Val Val Pro Ser Phe  
 660 665 670

Leu Pro Val Asp Gln Gly Gly Ser Leu Val Gly Arg Asn Gly Val Gly  
 675 680 685

Gly Met Ala Lys Glu Ala Thr Met Lys Gly Ser Ser Ser Ala Ser Ile  
 690 695 700

Val Lys Gly Gln His Glu Met Ser Glu Met Asp Gly Arg Trp Glu Glu  
 705 710 715 720

His Arg Ser Leu Leu Ser Gly Arg Ala Thr Gln Phe Thr Gly Ala Thr  
 725 730 735

Gly Ala Ile Met Thr Thr Glu Thr Thr Lys Thr Ala Arg Ala Thr Gly  
 740 745 750

Ala Ser Arg Asp Met Ala Gly Ala Gln Ala Ala Ala Val Ala Leu Asn  
 755 760 765

Glu Glu Phe Leu Arg Asn Tyr Phe Thr Asp Lys Ala Ala Ser Tyr Thr  
 770 775 780

Glu Glu Asp Glu Asn His Thr Ala Lys Asp Cys Leu Leu Val Tyr Ser  
 785 790 795 800

Gln Glu Glu Thr Glu Ser Leu Asn Ala Ser Ile Gly Cys Cys Ser Phe  
 805 810 815

Ile Glu Gly Glu Leu Asp Asp Arg Phe Leu Asp Asp Leu Gly Leu Lys  
 820 825 830

Phe Lys Thr Leu Ala Glu Val Cys Leu Gly Gln Lys Ile Asp Ile Asn  
 835 840 845



Lys Glu Ile Glu Gln Arg Gln Lys Pro Ala Thr Glu Thr Ser Met Asn  
 850 855 860

Thr Ala Ser His Ser Leu Cys Glu Gln Thr Met Val Asn Ser Glu Asn  
 865 870 875 880

Thr Tyr Ser Ser Gly Ser Ser Phe Pro Val Pro Lys Ser Leu Gln Glu  
 885 890 895

Ala Asn Ala Glu Lys Val Thr Gln Glu Ile Val Thr Glu Arg Ser Val  
 900 905 910

Ser Ser Arg Gln Ala Gln Lys Val Ala Thr Pro Leu Pro Asp Pro Met  
 915 920 925

Ala Ser Arg Asn Val Ile Ala Thr Glu Thr Ser Tyr Val Thr Gly Ser  
 930 935 940

Thr Met Pro Pro Thr Thr Val Ile Leu Gly Pro Ser Gln Pro Gln Ser  
 945 950 955 960

Leu Ile Val Thr Glu Arg Val Tyr Ala Pro Ala Ser Thr Leu Val Asp  
 965 970 975

Gln Pro Tyr Ala Asn Glu Gly Thr Val Val Val Thr Glu Arg Val Ile  
 980 985 990

Gln Pro His Gly Gly Gly Ser Asn Pro Leu Glu Gly Thr Gln His Leu  
 995 1000 1005

Gln Asp Val Pro Tyr Val Met Val Arg Glu Arg Glu Ser Phe Leu Ala  
 1010 1015 1020

Pro Ser Ser Gly Val Gln Pro Thr Leu Ala Met Pro Asn Ile Ala Val  
 1025 1030 1035 1040

Gly Gln Asn Val Thr Val Thr Glu Arg Val Leu Ala Pro Ala Ser Thr  
 1045 1050 1055

Leu Gln Ser Ser Tyr Gln Ile Pro Thr Glu Asn Ser Met Thr Ala Arg  
 1060 1065 1070

Asn Thr Thr Val Ser Gly Ala Gly Val Pro Gly Pro Leu Pro Asp Phe  
 1075 1080 1085

Gly Leu Glu Glu Ser Gly His Ser Asn Ser Thr Ile Thr Thr Ser Ser  
 1090 1095 1100





Thr Arg Val Thr Lys His Ser Thr Val Gln His Ser Tyr Ser  
 1105 1110 1115

<210> 43  
 <211> 97  
 <212> PRT  
 <213> Homo sapiens

<400> 43  
 Met Thr Lys Gly Thr Ser Ser Phe Gly Lys Arg Arg Asn Lys Thr His  
 1 5 10 15

Thr Leu Cys Arg Arg Cys Gly Ser Lys Ala Tyr His Leu Gln Lys Ser  
 20 25 30

Thr Cys Gly Lys Cys Gly Tyr Pro Ala Lys Arg Lys Arg Lys Tyr Asn  
 35 40 45

Trp Ser Ala Lys Ala Lys Arg Arg Asn Thr Thr Gly Thr Gly Arg Met  
 50 55 60

Arg His Leu Lys Ile Val Tyr Arg Arg Phe Arg His Gly Phe Arg Glu  
 65 70 75 80

Gly Thr Thr Pro Lys Pro Lys Arg Ala Ala Val Ala Ala Ser Ser Ser  
 85 90 95

Ser

<210> 44  
 <211> 889  
 <212> PRT  
 <213> Homo sapiens

<400> 44  
 Met Ala Ala Ala Val Gly Val Arg Gly Arg Tyr Glu Leu Pro Pro Cys  
 1 5 10 15

Ser Gly Pro Gly Trp Leu Leu Ser Leu Ser Ala Leu Leu Ser Val Ala  
 20 25 30

Ala Arg Gly Ala Phe Ala Thr Thr His Trp Val Val Thr Glu Asp Gly  
 35 40 45

Lys Ile Gln Gln Gln Val Asp Ser Pro Met Asn Leu Lys His Pro His



50	55	60
Asp Leu Val Ile Leu Met Arg Gln Glu Ala Thr Val Asn Tyr Leu Lys		
65	70	75 80
Glu Leu Glu Lys Gln Leu Val Ala Gln Lys Ile His Ile Glu Glu Asn		
85	90	95
Glu Asp Arg Asp Thr Gly Leu Glu Gln Arg His Asn Lys Glu Asp Pro		
100	105	110
Asp Cys Ile Lys Ala Lys Val Pro Leu Gly Asp Leu Asp Leu Tyr Asp		
115	120	125
Gly Thr Tyr Ile Thr Leu Glu Ser Lys Asp Ile Ser Pro Glu Asp Tyr		
130	135	140
Ile Asp Thr Glu Ser Pro Val Pro Pro Asp Pro Glu Gln Pro Asp Cys		
145	150	155 160
Thr Lys Ile Leu Glu Leu Pro Tyr Ser Ile His Ala Phe Gln His Leu		
165	170	175
Arg Gly Val Gln Glu Arg Val Asn Leu Ser Ala Pro Leu Leu Pro Lys		
180	185	190
Glu Asp Pro Ile Phe Thr Tyr Leu Ser Lys Arg Leu Gly Arg Ser Ile		
195	200	205
Asp Asp Ile Gly His Leu Ile His Glu Gly Leu Gln Lys Asn Thr Ser		
210	215	220
Ser Trp Val Leu Tyr Asn Met Ala Ser Phe Tyr Trp Arg Ile Lys Asn		
225	230	235 240
Glu Pro Tyr Gln Val Val Glu Cys Ala Met Arg Ala Leu His Phe Ser		
245	250	255
Ser Arg His Asn Lys Asp Ile Ala Leu Val Asn Leu Ala Asn Val Leu		
260	265	270
His Arg Ala His Phe Ser Ala Asp Ala Ala Val Val Val His Ala Ala		
275	280	285
Leu Asp Asp Ser Asp Phe Phe Thr Ser Tyr Tyr Thr Leu Gly Asn Ile		
290	295	300
Tyr Ala Met Leu Gly Glu Tyr Asn His Ser Val Leu Cys Tyr Asp His		



305	310	315	320
Ala Leu Gln Ala Arg Pro Gly Phe Glu Gln Ala Ile Lys Arg Lys His			
325	330	335	
Ala Val Leu Cys Gln Gln Lys Leu Glu Gln Lys Leu Glu Ala Gln His			
340	345	350	
Arg Ser Leu Gln Arg Thr Leu Asn Glu Leu Lys Glu Tyr Gln Lys Gln			
355	360	365	
His Asp His Tyr Leu Arg Gln Gln Glu Ile Leu Glu Lys His Lys Leu			
370	375	380	
Ile Gln Glu Glu Gln Ile Leu Arg Asn Ile Ile His Glu Thr Gln Met			
385	390	395	400
Ala Lys Glu Ala Gln Leu Gly Asn His Gln Ile Cys Arg Leu Val Asn			
405	410	415	
Gln Gln His Ser Leu His Cys Gln Trp Asp Gln Pro Val Arg Tyr His			
420	425	430	
Arg Gly Asp Ile Phe Glu Asn Val Asp Tyr Val Gln Phe Gly Glu Asp			
435	440	445	
Ser Ser Thr Ser Ser Met Met Ser Val Asn Phe Asp Val Gln Ser Asn			
450	455	460	
Gln Ser Asp Ile Asn Asp Ser Val Lys Ser Ser Pro Val Ala His Ser			
465	470	475	480
Ile Leu Trp Ile Trp Gly Arg Asp Ser Asp Ala Tyr Arg Asp Lys Gln			
485	490	495	
His Ile Leu Trp Pro Lys Arg Ala Asp Cys Thr Glu Ser Tyr Pro Arg			
500	505	510	
Val Pro Val Gly Gly Glu Leu Pro Thr Tyr Phe Leu Pro Pro Glu Asn			
515	520	525	
Lys Gly Leu Arg Ile His Glu Leu Ser Ser Asp Asp Tyr Ser Thr Glu			
530	535	540	
Glu Glu Ala Gln Thr Pro Asp Cys Ser Ile Thr Asp Phe Arg Lys Ser			
545	550	555	560
His Thr Leu Ser Tyr Leu Val Lys Glu Leu Glu Val Arg Met Asp Leu			



565					570					575						
Lys	Ala	Lys	Met	Pro	Asp	Asp	His	Ala	Arg	Lys	Ile	Leu	Leu	Ser	Arg	
580					585					590						
Ile	Asn	Asn	Tyr	Thr	Ile	Pro	Glu	Glu	Glu	Ile	Gly	Ser	Phe	Leu	Phe	
595					600					605						
His	Ala	Ile	Asn	Lys	Pro	Asn	Ala	Pro	Ile	Trp	Leu	Ile	Leu	Asn	Glu	
610					615					620						
Ala	Gly	Leu	Tyr	Trp	Arg	Ala	Val	Gly	Asn	Ser	Thr	Phe	Ala	Ile	Ala	
625					630					635					640	
Cys	Leu	Gln	Arg	Ala	Leu	Asn	Leu	Ala	Pro	Leu	Gln	Tyr	Gln	Asp	Val	
645					650					655						
Pro	Leu	Val	Asn	Leu	Ala	Asn	Leu	Leu	Ile	His	Tyr	Gly	Leu	His	Leu	
660					665					670						
Asp	Ala	Thr	Lys	Leu	Leu	Leu	Gln	Ala	Leu	Ala	Ile	Asn	Ser	Ser	Glu	
675					680					685						
Pro	Leu	Thr	Phe	Leu	Ser	Leu	Gly	Asn	Ala	Tyr	Leu	Ala	Leu	Lys	Asn	
690					695					700						
Ile	Ser	Gly	Ala	Leu	Glu	Ala	Phe	Arg	Gln	Ala	Leu	Lys	Leu	Thr	Thr	
705					710					715					720	
Lys	Cys	Pro	Glu	Cys	Glu	Asn	Ser	Leu	Lys	Leu	Ile	Arg	Cys	Met	Gln	
725					730					735						
Phe	Tyr	Pro	Phe	Leu	Tyr	Asn	Ile	Thr	Ser	Ser	Val	Cys	Ser	Gly	Thr	
740					745					750						
Val	Val	Glu	Glu	Ser	Asn	Gly	Ser	Asp	Glu	Met	Glu	Asn	Ser	Asp	Glu	
755					760					765						
Thr	Lys	Met	Ser	Glu	Glu	Ile	Leu	Ala	Leu	Val	Asp	Glu	Phe	Gln	Gln	
770					775					780						
Ala	Trp	Pro	Leu	Glu	Gly	Phe	Gly	Gly	Ala	Leu	Glu	Met	Lys	Gly	Arg	
785					790					795					800	
Arg	Leu	Asp	Leu	Gln	Gly	Ile	Arg	Val	Leu	Lys	Lys	Gly	Pro	Gln	Asp	
805					810					815						
Gly	Val	Ala	Arg	Ser	Ser	Cys	Tyr	Gly	Asp	Cys	Arg	Ser	Glu	Asp	Asp	





820				825				830							
Glu	Ala	Thr	Glu	Trp	Ile	Thr	Phe	Gln	Val	Lys	Arg	Val	Lys	Lys	Pro
835				840				845							
Lys	Gly	Asp	His	Lys	Lys	Thr	Pro	Gly	Lys	Lys	Val	Glu	Thr	Gly	Gln
850				855				860							
Ile	Glu	Asn	Gly	His	Arg	Tyr	Gln	Ala	Asn	Leu	Glu	Ile	Thr	Gly	Pro
865				870				875				880			
Lys	Val	Ala	Ser	Pro	Gly	Pro	Gln	Gly							
885															

```
<210> 45
<211> 690
<212> PRT
<213> Homo sapiens
```

```
<400> 45
Phe Leu Thr Leu Phe Ile Phe Arg Ser Gly Leu Cys Arg Gly Asn Ser
  1           5           10          15
```

Val Glu Arg Lys Ile Tyr Ile Pro Leu Asn Lys Thr Ala Pro Cys Val  
20 25 30

Arg Leu Leu Asn Ala Thr His Gln Ile Gly Cys Gln Ser Ser Ile Ser  
35 40 45

Gly Asp Thr Gly Val Ile His Val Val Glu Lys Glu Glu Asp Leu Gln  
50 55 60

Trp Val Leu Thr Asp Gly Pro Asn Pro Pro Tyr Met Val Leu Leu Glu  
65 70 75 80

Ser Lys His Phe Thr Arg Asp Leu Met Glu Lys Leu Lys Gly Arg Thr  
85 90 95

Ser Arg Ile Ala Gly Leu Ala Val Ser Leu Thr Lys Pro Ser Pro Ala  
100 105 110

Ser Gly Phe Ser Pro Ser Val Gln Cys Pro Asn Asp Gly Phe Gly Val  
115 120 125

Tyr Ser Asn Ser Tyr Gly Pro Glu Phe Ala His Cys Arg Glu Ile Gln  
130 135 140



Trp Asn Ser Leu Gly Asn Gly Leu Ala Tyr Glu Asp Phe Ser Phe Pro  
 145 150 155 160  
 Ile Phe Leu Leu Glu Asp Glu Asn Glu Thr Lys Val Ile Lys Gln Cys  
 165 170 175  
 Tyr Gln Asp His Asn Leu Ser Gln Asn Gly Ser Ala Pro Thr Phe Pro  
 180 185 190  
 Leu Cys Ala Met Gln Leu Phe Ser His Met His Ala Val Ile Ser Thr  
 195 200 205  
 Ala Thr Cys Met Arg Arg Ser Ser Ile Gln Ser Thr Phe Ser Ile Asn  
 210 215 220  
 Pro Glu Ile Val Cys Asp Pro Leu Ser Asp Tyr Asn Val Trp Ser Met  
 225 230 235 240  
 Leu Lys Pro Ile Asn Thr Thr Gly Thr Leu Lys Pro Asp Asp Arg Val  
 245 250 255  
 Val Val Ala Ala Thr Arg Leu Asp Ser Arg Ser Phe Phe Trp Asn Val  
 260 265 270  
 Ala Pro Gly Ala Glu Ser Ala Val Ala Ser Phe Val Thr Gln Leu Ala  
 275 280 285  
 Ala Ala Glu Ala Leu Gln Lys Ala Pro Asp Val Thr Thr Leu Pro Arg  
 290 295 300  
 Asn Val Met Phe Val Phe Phe Gln Gly Glu Thr Phe Asp Tyr Ile Gly  
 305 310 315 320  
 Ser Ser Arg Met Val Tyr Asp Met Glu Lys Gly Lys Phe Pro Val Gln  
 325 330 335  
 Leu Glu Asn Val Asp Ser Phe Val Glu Leu Gly Gln Val Ala Leu Arg  
 340 345 350  
 Thr Ser Leu Glu Leu Trp Met His Thr Asp Pro Val Ser Gln Lys Asn  
 355 360 365  
 Glu Ser Val Arg Asn Gln Val Glu Asp Leu Leu Ala Thr Leu Glu Lys  
 370 375 380  
 Ser Gly Ala Gly Val Pro Ala Val Ile Leu Arg Arg Pro Asn Gln Ser  
 385 390 395 400



Gln Pro Leu Pro Pro Ser Ser Leu Gln Arg Phe Leu Arg Ala Arg Asn  
 405 410 415  
 Ile Ser Gly Val Val Leu Ala Asp His Ser Gly Ala Phe His Asn Lys  
 420 425 430  
 Tyr Tyr Gln Ser Ile Tyr Asp Thr Ala Glu Asn Ile Asn Val Ser Tyr  
 435 440 445  
 Pro Glu Trp Leu Ser Pro Glu Glu Asp Leu Asn Phe Val Thr Asp Thr  
 450 455 460  
 Ala Lys Ala Leu Ala Asp Val Ala Thr Val Leu Gly Arg Ala Leu Tyr  
 465 470 475 480  
 Glu Leu Ala Gly Gly Thr Asn Phe Ser Asp Thr Val Gln Ala Asp Pro  
 485 490 495  
 Gln Thr Val Thr Arg Leu Leu Tyr Gly Phe Leu Ile Lys Ala Asn Asn  
 500 505 510  
 Ser Trp Phe Gln Ser Ile Leu Arg Gln Asp Leu Arg Ser Tyr Leu Gly  
 515 520 525  
 Asp Gly Pro Leu Gln His Tyr Ile Ala Val Ser Ser Pro Thr Asn Thr  
 530 535 540  
 Thr Tyr Val Val Gln Tyr Ala Leu Ala Asn Leu Thr Gly Thr Val Val  
 545 550 555 560  
 Asn Leu Thr Arg Glu Gln Cys Gln Asp Pro Ser Lys Val Pro Ser Glu  
 565 570 575  
 Asn Lys Asp Leu Tyr Glu Tyr Ser Trp Val Gln Gly Pro Leu His Ser  
 580 585 590  
 Asn Glu Thr Asp Arg Leu Pro Arg Cys Val Arg Ser Thr Ala Arg Leu  
 595 600 605  
 Ala Arg Ala Leu Ser Pro Ala Phe Glu Leu Ser Gln Trp Ser Ser Thr  
 610 615 620  
 Glu Tyr Ser Thr Trp Thr Glu Ser Arg Trp Lys Asp Ile Arg Ala Arg  
 625 630 635 640  
 Ile Phe Leu Ile Ala Ser Lys Glu Leu Glu Leu Ile Thr Leu Thr Val  
 645 650 655



Gly Phe Gly Ile Leu Ile Phe Ser Leu Ile Val Thr Tyr Cys Ile Asn  
                   660                                  665                                  670

Ala Lys Ala Asp Val Leu Phe Ile Ala Pro Arg Glu Pro Gly Ala Val  
                   675                                  680                                  685

Ser Tyr  
           690

<210> 46  
 <211> 170  
 <212> PRT  
 <213> Homo sapiens

<400> 46  
 Gln Val Pro Arg Ser Lys Ala Leu Glu Val Thr Lys Leu Ala Ile Glu  
   1                                  5                                  10                                  15

Ala Gly Phe Arg His Ile Asp Ser Ala His Leu Tyr Asn Asn Glu Glu  
                                   20                                  25                                  30

Gln Val Gly Leu Ala Ile Arg Ser Lys Ile Ala Asp Gly Ser Val Lys  
                   35                                  40                                  45

Arg Glu Asp Ile Phe Tyr Thr Ser Lys Leu Trp Ser Thr Phe His Arg  
           50                                  55                                  60

Pro Glu Leu Val Arg Pro Ala Leu Glu Asn Ser Leu Lys Lys Ala Gln  
   65                                  70                                  75                                  80

Leu Asp Tyr Val Asp Leu Tyr Leu Ile His Ser Pro Met Ser Leu Lys  
                                   85                                  90                                  95

Pro Gly Glu Glu Leu Ser Pro Thr Asp Glu Gln Val Ala Lys Val Ile  
                   100                                  105                                  110

Phe Asp Ile Val Asp Leu Cys Thr Thr Trp Glu Gly Met Glu Lys Cys  
           115                                  120                                  125

Lys Asp Gly Arg Asn Trp Gly Lys Ser Ile Gly Val Ser His Phe Asn  
   130                                  135                                  140

Pro Gln Ala Leu Gly Met Ser Leu Gln Lys Ala Gly Ile Gln Leu Lys  
   145                                  150                                  155                                  160

Arg Ser Ala Pro Val Glu Cys Pro Ile Tyr  
                                   165                                  170





&lt;210&gt; 47

&lt;211&gt; 1596

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 47

Met Thr Thr Glu Thr Gly Pro Asp Ser Glu Val Lys Lys Ala Gln Glu  
 1 5 10 15

Glu Ala Pro Gln Gln Pro Glu Ala Ala Ala Val Thr Thr Pro Val  
 20 25 30

Thr Pro Ala Gly His Gly His Pro Glu Ala Asn Ser Asn Glu Lys His  
 35 40 45

Pro Ser Gln Gln Asp Thr Arg Pro Ala Glu Gln Ser Leu Asp Met Glu  
 50 55 60

Glu Lys Asp Tyr Ser Glu Ala Asp Gly Leu Ser Glu Arg Thr Thr Pro  
 65 70 75 80

Ser Lys Ala Gln Lys Ser Pro Gln Lys Ile Ala Lys Lys Tyr Lys Ser  
 85 90 95

Ala Ile Cys Arg Val Thr Leu Leu Asp Ala Ser Glu Tyr Glu Cys Glu  
 100 105 110

Val Glu Lys His Gly Arg Gly Gln Val Leu Phe Asp Leu Val Cys Glu  
 115 120 125

His Leu Asn Leu Leu Glu Lys Asp Tyr Phe Gly Leu Thr Phe Cys Asp  
 130 135 140

Ala Asp Ser Gln Lys Asn Trp Leu Asp Pro Ser Lys Glu Ile Lys Lys  
 145 150 155 160

Gln Ile Arg Ser Glu Trp Leu Val Val Phe Gly Glu Val Gly Ser Pro  
 165 170 175

Trp Asn Phe Ala Phe Thr Val Lys Phe Tyr Pro Pro Asp Pro Ala Gln  
 180 185 190

Leu Thr Glu Asp Ile Thr Arg Tyr Tyr Leu Cys Leu Gln Leu Arg Ala  
 195 200 205

Asp Ile Ile Thr Gly Arg Leu Pro Cys Ser Phe Val Thr His Ala Leu



210	215	220
Leu Gly Ser Tyr Ala Val Gln Ala Glu Leu Gly Asp Tyr Asp Ala Glu		
225	230	235 240
Glu His Val Gly Asn Tyr Val Ser Glu Leu Arg Phe Ala Pro Asn Gln		
245	250	255
Thr Arg Glu Leu Glu Glu Arg Ile Met Glu Leu His Lys Thr Tyr Arg		
260	265	270
Gly Met Thr Pro Gly Glu Ala Glu Ile His Phe Leu Glu Asn Ala Lys		
275	280	285
Lys Leu Ser Met Tyr Gly Val Asp Leu His His Ala Lys Asp Ser Glu		
290	295	300
Gly Ile Asp Ile Met Leu Gly Val Cys Ala Asn Gly Leu Leu Ile Tyr		
305	310	315 320
Arg Asp Arg Leu Arg Ile Asn Arg Phe Ala Trp Pro Lys Ile Leu Lys		
325	330	335
Ile Ser Tyr Lys Arg Ser Asn Phe Tyr Ile Lys Ile Arg Pro Gly Glu		
340	345	350
Tyr Glu Gln Phe Glu Ser Thr Ile Gly Phe Lys Leu Pro Asn His Arg		
355	360	365
Ser Ala Lys Arg Leu Trp Lys Val Cys Ile Glu His His Thr Phe Phe		
370	375	380
Arg Leu Val Ser Pro Glu Pro Pro Pro Lys Gly Phe Leu Val Met Gly		
385	390	395 400
Ser Lys Phe Arg Tyr Ser Gly Arg Thr Gln Ala Gln Thr Arg Gln Ala		
405	410	415
Ser Ala Leu Ile Asp Arg Pro Ala Pro Phe Phe Glu Arg Ser Ser Ser		
420	425	430
Lys Arg Tyr Thr Met Ser Arg Ser Leu Asp Gly Ala Glu Phe Ser Arg		
435	440	445
Pro Ala Ser Val Ser Glu Asn His Asp Ala Gly Pro Asp Gly Asp Lys		
450	455	460
Arg Asp Glu Asp Gly Glu Ser Gly Gly Gln Arg Ser Glu Ala Glu Glu		



465	470	475	480
Gly Glu Val Arg Thr Pro Thr Lys Ile Lys Glu Leu Lys Phe Leu Asp			
485	490	495	
Lys Pro Glu Asp Val Leu Leu Lys His Gln Ala Ser Ile Asn Glu Leu			
500	505	510	
Lys Arg Thr Leu Lys Glu Pro Asn Ser Lys Leu Ile His Arg Asp Arg			
515	520	525	
Asp Trp Glu Arg Glu Arg Arg Leu Pro Ser Ser Pro Ala Ser Pro Ser			
530	535	540	
Pro Lys Gly Thr Pro Glu Lys Ala Asn Glu Ser Gln Arg Thr Gln Asp			
545	550	555	560
Ile Ser Gln Arg Asp Leu Val Pro Glu Pro Gly Ala Ala Ala Gly Leu			
565	570	575	
Glu Val Phe Thr Gln Lys Ser Leu Ala Ala Ser Pro Glu Gly Ser Glu			
580	585	590	
His Trp Val Phe Ile Glu Arg Glu Tyr Thr Arg Pro Glu Glu Leu Gly			
595	600	605	
Leu Leu Lys Val Thr Thr Met Gln Gln Glu Glu Arg Gln Ala Gly Leu			
610	615	620	
Ala Gly Ile Leu Ala Asn Gly Arg Leu Ser Lys Val Asp Val Leu Val			
625	630	635	640
Asp Lys Phe Lys Val Glu Val Ala Thr Glu Glu Met Val Gly Asn Arg			
645	650	655	
Arg Ala Asn Thr Gln Gln Gln Gly Lys Met Ile Ala Ser Pro Glu Asp			
660	665	670	
Phe Glu Ser Val Gly Glu Glu Gly Pro Trp Ile Arg Glu Ser Pro Gly			
675	680	685	
Gly Ala Ala Leu Ala Ser Gly Arg Thr Leu Ala Glu Lys Leu Leu Glu			
690	695	700	
Gly Ser Glu Leu Arg Ala Asp Thr Arg Glu Ala Thr Ile Arg Asn Arg			
705	710	715	720
Cys Met Ser Asp Gly Gln Pro Glu Gly Gln Thr Glu Leu Arg Lys Gly			



725	730	735
Leu Glu Glu Pro His Thr Cys Gly Arg Pro Thr Ala Pro Gly Thr Arg 740	745	750
Pro Ala Glu Val Asp Val Leu Ser Pro Ala Ser Asp Lys Gly Gly Leu 755	760	765
Gln Ser Phe Leu Leu Asp Pro Ala His Ala Glu Ala Arg Ala Glu Leu 770	775	780
Ser Asn Glu Thr Asp Thr Ser Phe Ala Glu Arg Ser Phe Tyr Leu Asn 785	790	795
Tyr Glu Glu Lys Asp Ser Glu Asp Gln Val Leu Pro Pro Pro Leu Glu 805	810	815
Glu Arg Lys Gly Arg Leu Asp Ala Pro Pro Gly Gly Glu Pro Arg Pro 820	825	830
Thr Leu Asn Ser Leu Asp Leu Arg Val Ser Ala Ala Ala Ser Ser Arg 835	840	845
Ser Lys Asp Glu Ala His Met Thr Ser Pro Lys Glu Gly Ala Gly Thr 850	855	860
Pro Lys Asn His Gly Gly Pro Gly Asp Leu Lys Gly Ser Pro Ala Gly 865	870	875
Gln Thr Phe Ala Glu Gly Trp Glu Asp Ala Gln Trp Gly Val Glu Gly 885	890	895
Glu Phe Pro His Leu Thr Ala Ser Ala Ala Arg Glu Glu Gly Thr Pro 900	905	910
Val Ser Gly Asp Leu Leu Gly Lys Ala Glu Glu Ser Pro Thr Glu Glu 915	920	925
Leu Lys Lys His Pro Pro His Arg Gly Gln Gly Val His Pro Asp Pro 930	935	940
Gln Ala Cys Ala Leu Pro Arg Ala Ile Pro Leu Asn Val Arg Lys Pro 945	950	955
Val Lys Pro Asp Arg Gly Asn Phe Pro Pro Lys Glu Arg Gly Val Val 965	970	975
Pro Thr Gln Lys Gly Gly Ala Glu Leu Lys Asp Arg Glu Ala Ser Ala		





980	985	990
Phe Leu His Met Glu Val Ile Ile Pro Leu Pro Ala Ser Pro Gly His		
995	1000	1005
Ser Glu Asp Leu Ala Ala Leu Glu Glu Ala Ser Pro Ser Pro Thr Ser		
1010	1015	1020
His Gly Ser Gly Glu Pro Ser Glu Leu Arg Glu Pro Phe Leu Arg His		
1025	1030	1035 1040
Val His Leu Ser Lys Ala Ser Pro Glu Pro Lys Asp Gln Val Gly Phe		
1045	1050	1055
Val Val Ser Pro Ala Thr Gly Gly Glu Arg Arg Pro Pro Pro Ile Thr		
1060	1065	1070
Ser Arg Lys Pro Arg Val Val Pro Glu Glu Ala Glu Gly Arg Ile Pro		
1075	1080	1085
Leu Gly Phe Gly Phe Pro Ser Gly Lys Arg Arg Glu Met Thr Ser Phe		
1090	1095	1100
Gln Ala Gly Asp Gln Glu Gly Ser Leu Glu Asp Ile Ser Lys Thr Ser		
1105	1110	1115 1120
Val Ala Asn Lys Ile Arg Ile Phe Glu Thr His Gly Ala Glu Thr Arg		
1125	1130	1135
Arg Met Ser Glu Gly Glu Ala Arg Ser Leu Pro Asn Asp Val Ser Ser		
1140	1145	1150
Glu Ala Pro Val Gly Gln Ala Glu Gln Gln Arg Ser Thr Leu Ser Asp		
1155	1160	1165
Leu Gly Phe Ala Gln Leu Gln Pro Pro Gly Asp Phe Ala Ser Pro Lys		
1170	1175	1180
Ala Thr His Ser Thr Val Ile Pro Leu Ala Thr Arg His Phe Arg Glu		
1185	1190	1195 1200
Asp Thr Ser Ala Ser Tyr Gln Glu Ala His Thr Glu Leu Glu Pro Val		
1205	1210	1215
Ser Pro Asn Ser Gly Cys Glu Thr Thr Leu Ala Glu Ala Thr Gly Thr		
1220	1225	1230
Gly Val Thr Gly Arg Asn Lys Ser Gly Asp Ala Val Arg Glu Glu Lys		



1235	1240	1245
Arg Ser Thr Asn Leu Ala Ala Asn Thr Pro Gly Lys Gly Arg Leu 1250	1255	1260
Arg Phe Ala Ser Pro Ser Gly Pro Gln Arg Ala Gly Leu Arg Glu Gly 1265	1270	1275 1280
Ser Glu Glu Lys Val Lys Pro Pro Arg Pro Arg Ala Pro Glu Ser Asp 1285	1290	1295
Thr Gly Asp Glu Asp Gln Asp Gln Glu Arg Asp Thr Val Phe Leu Lys 1300	1305	1310
Asp Asn His Leu Ala Ile Glu Arg Lys Cys Ser Ser Ile Thr Val Ser 1315	1320	1325
Ser Thr Ser Ser Leu Glu Ala Glu Val Asp Phe Thr Val Ile Gly Asp 1330	1335	1340
Tyr His Gly Ser Ala Phe Glu Asp Phe Ser Arg Ser Leu Pro Glu Leu 1345	1350	1355 1360
Asp Arg Asp Lys Ser Asp Ser Asp Thr Glu Gly Leu Leu Phe Ser Arg 1365	1370	1375
Asp Leu Asn Lys Gly Ala Pro Ser Gln Asp Asp Glu Ser Gly Gly Ile 1380	1385	1390
Glu Asp Ser Pro Asp Arg Gly Ala Cys Ser Thr Pro Asp Met Pro Gln 1395	1400	1405
Phe Glu Pro Val Lys Thr Glu Thr Met Thr Val Ser Ser Leu Ala Ile 1410	1415	1420
Arg Lys Lys Ile Glu Pro Glu Ala Val Leu Gln Thr Arg Val Ser Ala 1425	1430	1435 1440
Met Asp Asn Thr Gln Val Asp Gly Ser Ala Ser Val Gly Arg Glu Phe 1445	1450	1455
Ile Ala Thr Thr Pro Ser Ile Thr Thr Glu Thr Ile Ser Thr Thr Met 1460	1465	1470
Glu Asn Ser Leu Lys Ser Gly Lys Gly Ala Ala Ala Met Ile Pro Gly 1475	1480	1485
Pro Gln Thr Val Ala Thr Glu Ile Arg Ser Leu Ser Pro Ile Ile Gly		



1490                      1495                      1500  
 Lys Asp Val Leu Thr Ser Thr Tyr Gly Ala Thr Ala Glu Thr Leu Ser  
 1505                      1510                      1515                      1520  
 Thr Ser Thr Thr Thr His Val Thr Lys Thr Val Lys Gly Gly Phe Ser  
                     1525                      1530                      1535  
 Glu Thr Arg Ile Glu Lys Arg Ile Ile Ile Thr Gly Asp Glu Asp Val  
                     1540                      1545                      1550  
 Asp Gln Asp Gln Ala Leu Ala Leu Ala Ile Lys Glu Ala Lys Leu Gln  
                     1555                      1560                      1565  
 His Pro Asp Met Leu Val Thr Lys Ala Val Val Tyr Arg Glu Thr Asp  
                     1570                      1575                      1580  
 Pro Ser Pro Glu Glu Arg Asp Lys Lys Pro Gln Lys  
 1585                      1590                      1595  
  
 <210> 48  
 <211> 455  
 <212> PRT  
 <213> Homo sapiens  
  
 <400> 48  
 Met Ala Ala Pro Glu Glu His Asp Ser Pro Thr Glu Ala Ser Gln Pro  
   1                      5                      10                      15  
 Ile Val Glu Glu Glu Glu Thr Lys Thr Phe Lys Asp Leu Gly Val Thr  
                     20                      25                      30  
 Asp Val Leu Cys Glu Ala Cys Asp Gln Leu Gly Trp Thr Lys Pro Thr  
                     35                      40                      45  
 Lys Ile Gln Ile Glu Ala Ile Pro Leu Ala Leu Gln Gly Arg Asp Ile  
                     50                      55                      60  
 Ile Gly Leu Ala Glu Thr Gly Ser Gly Lys Thr Gly Ala Phe Ala Leu  
   65                      70                      75                      80  
 Pro Ile Leu Asn Ala Leu Leu Glu Thr Pro Gln Arg Leu Phe Ala Leu  
                     85                      90                      95  
 Val Leu Thr Pro Thr Arg Glu Leu Ala Phe Gln Ile Ser Glu Gln Phe  
                     100                      105                      110



Glu Ala Leu Gly Ser Ser Ile Gly Val Gln Ser Ala Val Ile Val Gly  
 115 120 125

Gly Ile Asp Ser Met Ser Gln Ser Leu Ala Leu Ala Lys Lys Pro His  
 130 135 140

Ile Ile Ile Ala Thr Pro Gly Arg Leu Ile Asp His Leu Glu Asn Thr  
 145 150 155 160

Lys Gly Phe Asn Leu Arg Ala Leu Lys Tyr Leu Val Met Asp Glu Ala  
 165 170 175

Asp Arg Ile Leu Asn Met Asp Phe Glu Thr Glu Val Asp Lys Ile Leu  
 180 185 190

Lys Val Ile Pro Arg Asp Arg Lys Thr Phe Leu Phe Ser Ala Thr Met  
 195 200 205

Thr Lys Lys Val Gln Lys Leu Gln Arg Ala Ala Leu Lys Asn Pro Val  
 210 215 220

Lys Cys Ala Val Ser Ser Lys Tyr Gln Thr Val Glu Lys Leu Gln Gln  
 225 230 235 240

Tyr Tyr Ile Phe Ile Pro Ser Lys Phe Lys Asp Thr Tyr Leu Val Tyr  
 245 250 255

Ile Leu Asn Glu Leu Ala Gly Asn Ser Phe Met Ile Phe Cys Ser Thr  
 260 265 270

Cys Asn Asn Thr Gln Arg Thr Ala Leu Leu Leu Arg Asn Leu Gly Phe  
 275 280 285

Thr Ala Ile Pro Leu His Gly Gln Met Ser Gln Ser Lys Arg Leu Gly  
 290 295 300

Ser Leu Asn Lys Phe Lys Ala Lys Ala Arg Ser Ile Leu Leu Ala Thr  
 305 310 315 320

Asp Val Ala Ser Arg Gly Leu Asp Ile Pro His Val Asp Val Val Val  
 325 330 335

Asn Phe Asp Ile Pro Thr His Ser Lys Asp Tyr Ile His Arg Val Gly  
 340 345 350

Arg Thr Ala Arg Ala Gly Arg Ser Gly Lys Ala Ile Thr Phe Val Thr  
 355 360 365





Gln Tyr Asp Val Glu Leu Phe Gln Arg Ile Glu His Leu Ile Gly Lys  
 370 375 380

Lys Leu Pro Gly Phe Pro Thr Gln Asp Asp Glu Val Met Met Leu Thr  
 385 390 395 400

Glu Arg Val Ala Glu Ala Gln Arg Phe Ala Arg Met Glu Leu Arg Glu  
 405 410 415

His Gly Glu Lys Lys Lys Arg Ser Arg Glu Asp Ala Gly Asp Asn Asp  
 420 425 430

Asp Thr Glu Gly Ala Ile Gly Val Arg Asn Lys Val Ala Gly Gly Lys  
 435 440 445

Met Lys Lys Arg Lys Gly Arg  
 450 455

<210> 49

<211> 246

<212> PRT

<213> Homo sapiens

<400> 49

Met Ala Trp Ala Pro Leu Leu Leu Thr Leu Leu Ser Leu Leu Thr Gly  
 1 5 10 15

Ser Leu Ser Gln Pro Ile Leu Thr Gln Pro Pro Ser Ala Ser Ala Ser  
 20 25 30

Leu Gly Ala Ser Val Thr Leu Thr Cys Ser Val Ser Ser Asp Tyr Lys  
 35 40 45

Asn Leu Glu Val Asp Trp Phe Gln Gln Arg Pro Gly Lys Gly Pro Arg  
 50 55 60

Phe Val Met Arg Val Gly Thr Gly Gly Val Val Gly Phe Arg Gly Ala  
 65 70 75 80

Asp Ile Pro Asp Arg Phe Ser Val Ser Gly Ser Gly Leu Asn Arg Phe  
 85 90 95

Leu Thr Ile Arg Asn Ile Glu Glu Glu Asp Glu Ser Asp Tyr His Cys  
 100 105 110

Gly Thr Asp Leu Gly Ser Gly Thr Ser Phe Val Ser Trp Val Phe Gly  
 115 120 125



Gly Gly Thr Lys Leu Thr Val Leu Ser Gln Pro Lys Ala Ala Pro Ser  
 130 135 140

Val Thr Leu Phe Pro Pro Ser Ser Glu Glu Leu Gln Ala Asn Lys Ala  
 145 150 155 160

Thr Leu Val Cys Leu Ile Ser Asp Phe Tyr Pro Gly Ala Val Thr Val  
 165 170 175

Ala Trp Lys Ala Asp Ser Ser Pro Val Lys Ala Gly Val Glu Thr Thr  
 180 185 190

Thr Pro Ser Lys Gln Ser Asn Asn Lys Tyr Ala Ala Ser Ser Tyr Leu  
 195 200 205

Ser Leu Thr Pro Glu Gln Trp Lys Ser Asn Arg Ser Tyr Ser Cys Gln  
 210 215 220

Val Thr His Glu Gly Ser Thr Val Glu Lys Thr Val Ala Pro Thr Glu  
 225 230 235 240

Cys Ser Thr Glu Cys Ser  
 245

<210> 50

<211> 228

<212> PRT

<213> Homo sapiens

<400> 50

Ala Asn Ala Leu Gly Pro Cys Ala Glu Ile Val Met Thr Gln Thr Pro  
 1 5 10 15

Leu Ser Leu Ser Ile Thr Pro Gly Glu Gln Ala Ser Met Ser Cys Arg  
 20 25 30

Ser Ser Gln Ser Leu Leu His Ser Asp Gly Tyr Thr Tyr Leu Tyr Trp  
 35 40 45

Phe Leu Gln Lys Pro Gly Gln Ser Pro Gln Leu Leu Ile Tyr Glu Val  
 50 55 60

Ser Asn Arg Phe Ser Gly Val Ser Pro Ile Arg Phe Ser Gly Ser Gly  
 65 70 75 80

Ser Gly Arg Glu Phe Thr Leu Arg Ile Ser Arg Val Glu Ala Asp Asp



[illegible]

```
<210> 51
<211> 106
<212> PRT
<213> Homo sapiens
```

```

<400> 51
Gly Gln Pro Lys Ala Asn Pro Thr Val Thr Leu Phe Pro Pro Ser Ser
  1                      5                      10                      15
Glu Glu Leu Gln Ala Asn Lys Ala Thr Leu Val Cys Leu Ile Ser Asp
      20                      25                      30
Phe Tyr Pro Gly Ala Val Thr Val Ala Trp Lys Ala Asp Gly Ser Pro
      35                      40                      45
Val Lys Ala Gly Val Glu Thr Thr Lys Pro Ser Lys Gln Ser Asn Asn
      50                      55                      60

```



Lys Tyr Ala Ala Ser Ser Tyr Leu Ser Leu Thr Pro Glu Gln Trp Lys  
 65 70 75 80

Ser His Arg Ser Tyr Ser Cys Gln Val Thr His Glu Gly Ser Thr Val  
 85 90 95

Glu Lys Thr Val Ala Pro Thr Glu Cys Ser  
 100 105

<210> 52

<211> 56

<212> PRT

<213> Homo sapiens

<400> 52

Arg Thr Gly Tyr Glu Glu Glu Thr Trp Asn Leu Lys Glu Cys Val Gly  
 1 5 10 15

Arg Cys Ala Asn Pro Asn Val Asn Phe Leu Thr Lys Val Glu Ser Pro  
 20 25 30

Gly Met Val Gln Arg Trp Gly Leu Leu Leu Cys Arg Arg Asp Ser Arg  
 35 40 45

Phe Thr Pro Trp Gln Lys Ile Tyr  
 50 55

<210> 53

<211> 824

<212> PRT

<213> Homo sapiens

<400> 53

Met Ala Phe Ala Ser Phe Arg Arg Ile Leu Ala Leu Ser Thr Phe Glu  
 1 5 10 15

Lys Arg Lys Ser Arg Glu Tyr Glu His Val Arg Arg Asp Leu Asp Pro  
 20 25 30

Asn Glu Val Trp Glu Ile Val Gly Glu Leu Gly Asp Gly Ser Phe Gly  
 35 40 45

Met Val Tyr Lys Ala Lys Asn Lys Glu Thr Gly Ala Leu Ala Ala Ala  
 50 55 60

Ile Val Ile Glu Thr Lys Ser Glu Glu Glu Leu Glu Asp Tyr Ile Val





65	70	75	80
Glu Ile Glu Ile Leu Ala Thr Cys Asp His Pro Tyr Ile Val Lys Leu	85	90	95
Leu Gly Ala Tyr Tyr His Asp Gly Lys Leu Trp Ile Met Ile Glu Phe	100	105	110
Cys Pro Gly Gly Ala Val Asp Ala Ile Met Leu Glu Leu Asp Arg Gly	115	120	125
Leu Thr Glu Pro Gln Ile Gln Val Val Cys Arg Gln Met Leu Glu Ala	130	135	140
Leu Asn Phe Leu His Ser Lys Arg Ile Ile His Arg Asp Leu Lys Ala	145	150	155
Gly Asn Val Leu Met Thr Leu Glu Gly Asp Ile Arg Leu Ala Asp Phe	165	170	175
Gly Val Ser Ala Lys Asn Leu Lys Thr Leu Gln Lys Arg Asp Ser Phe	180	185	190
Ile Gly Thr Pro Tyr Trp Met Ala Pro Glu Val Val Met Cys Glu Thr	195	200	205
Met Lys Asp Thr Pro Tyr Asp Tyr Lys Ala Asp Ile Trp Ser Leu Gly	210	215	220
Ile Thr Leu Ile Glu Met Ala Gln Ile Glu Pro Pro His His Glu Leu	225	230	235
Asn Pro Met Arg Val Leu Leu Lys Ile Ala Lys Ser Asp Pro Pro Thr	245	250	255
Leu Leu Thr Pro Ser Lys Trp Ser Val Glu Phe Arg Asp Phe Leu Lys	260	265	270
Ile Ala Leu Asp Lys Asn Pro Glu Thr Arg Pro Ser Ala Ala Ala Ala	275	280	285
Leu Glu His Pro Phe Val Ser Ser Ile Thr Ser Asn Lys Ala Leu Arg	290	295	300
Glu Leu Val Ala Glu Ala Lys Ala Glu Val Met Glu Glu Ile Glu Asp	305	310	315
Gly Arg Asp Glu Gly Glu Glu Glu Asp Ala Val Asp Ala Ala Ser Thr			



325	330	335
Leu Glu Asn His Thr Gln Asn Ser Ser Glu Val Ser Pro Pro Ser Leu		
340	345	350
Asn Ala Asp Lys Pro Leu Glu Glu Ser Pro Ser Thr Pro Leu Ala Pro		
355	360	365
Ser Gln Ser Gln Asp Ser Val Asn Glu Pro Cys Ser Gln Pro Ser Gly		
370	375	380
Asp Arg Ser Leu Gln Thr Thr Ser Pro Pro Val Val Ala Pro Gly Asn		
385	390	395
Glu Asn Gly Leu Ala Val Pro Val Pro Leu Arg Lys Ser Arg Pro Val		
405	410	415
Ser Met Asp Ala Arg Ile Gln Val Ala Gln Glu Lys Gln Val Ala Glu		
420	425	430
Gln Gly Gly Asp Leu Ser Pro Ala Ala Asn Arg Ser Gln Lys Ala Ser		
435	440	445
Gln Ser Arg Pro Asn Ser Ser Ala Leu Glu Thr Leu Gly Gly Glu Lys		
450	455	460
Leu Ala Asn Gly Ser Leu Glu Pro Pro Ala Gln Ala Ala Pro Gly Pro		
465	470	475
Ser Lys Arg Asp Ser Asp Cys Ser Ser Leu Cys Thr Ser Glu Ser Met		
485	490	495
Asp Tyr Gly Thr Asn Leu Ser Thr Asp Leu Ser Leu Asn Lys Glu Met		
500	505	510
Gly Ser Leu Ser Ile Lys Asp Pro Lys Leu Tyr Lys Lys Thr Leu Lys		
515	520	525
Arg Thr Arg Lys Phe Val Val Asp Gly Val Glu Val Ser Ile Thr Thr		
530	535	540
Ser Lys Ile Ile Ser Glu Asp Glu Lys Lys Asp Glu Glu Met Arg Phe		
545	550	555
Leu Arg Arg Gln Glu Leu Arg Glu Leu Arg Leu Leu Gln Lys Glu Glu		
565	570	575
His Arg Asn Gln Thr Gln Leu Ser Asn Lys His Glu Leu Gln Leu Glu		



580	585	590
Gln Met His Lys Arg Phe Glu Gln Glu Ile Asn Ala Lys Lys Lys Phe 595	600	605
Phe Asp Thr Glu Leu Glu Asn Leu Glu Arg Gln Gln Lys Gln Gln Val 610	615	620
Glu Lys Met Glu Gln Asp His Ala Val Arg Arg Arg Glu Glu Ala Arg 625	630	635
Arg Ile Arg Leu Glu Gln Asp Arg Asp Tyr Thr Arg Phe Gln Glu Gln 645	650	655
Leu Lys Leu Met Lys Lys Glu Val Lys Asn Glu Val Glu Lys Leu Pro 660	665	670
Arg Gln Gln Arg Lys Glu Ser Met Lys Gln Lys Met Glu Glu His Thr 675	680	685
Gln Lys Lys Gln Leu Leu Asp Arg Asp Phe Val Ala Lys Gln Lys Glu 690	695	700
Asp Leu Glu Leu Ala Met Lys Arg Leu Thr Thr Asp Asn Arg Arg Glu 705	710	715
Ile Cys Asp Lys Glu Arg Glu Cys Leu Met Lys Lys Gln Glu Leu Leu 725	730	735
Arg Asp Arg Glu Ala Ala Leu Trp Glu Met Glu Glu His Gln Leu Gln 740	745	750
Glu Arg His Gln Leu Val Lys Gln Gln Leu Lys Asp Gln Tyr Phe Leu 755	760	765
Gln Arg His Glu Leu Leu Arg Lys His Glu Lys Glu Arg Glu Gln Met 770	775	780
Gln Arg Tyr Asn Gln Arg Met Ile Glu Gln Leu Lys Val Arg Gln Gln 785	790	795
Gln Glu Lys Ala Arg Leu Pro Lys Ile Gln Arg Ser Glu Gly Lys Thr 805	810	815
Arg Met Ala Met Tyr Lys Lys Ser 820		



&lt;210&gt; 54

&lt;211&gt; 1997

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 54

Met Leu Ser His Gly Ala Gly Leu Ala Leu Trp Ile Thr Leu Ser Leu  
 1 5 10 15

Leu Gln Thr Gly Leu Ala Glu Pro Glu Arg Cys Asn Phe Thr Leu Ala  
 20 25 30

Glu Ser Lys Ala Ser Ser His Ser Val Ser Ile Gln Trp Arg Ile Leu  
 35 40 45

Gly Ser Pro Cys Asn Phe Ser Leu Ile Tyr Ser Ser Asp Thr Leu Gly  
 50 55 60

Ala Ala Leu Cys Pro Thr Phe Arg Ile Asp Asn Thr Thr Tyr Gly Cys  
 65 70 75 80

Asn Leu Gln Asp Leu Gln Ala Gly Thr Ile Tyr Asn Phe Arg Ile Ile  
 85 90 95

Ser Leu Asp Glu Glu Arg Thr Val Val Leu Gln Thr Asp Pro Leu Pro  
 100 105 110

Pro Ala Arg Phe Gly Val Ser Lys Glu Lys Thr Thr Ser Thr Ser Leu  
 115 120 125

His Val Trp Trp Thr Pro Ser Ser Gly Lys Val Thr Ser Tyr Glu Val  
 130 135 140

Gln Leu Phe Asp Glu Asn Asn Gln Lys Ile Gln Gly Val Gln Ile Gln  
 145 150 155 160

Glu Ser Thr Ser Trp Asn Glu Tyr Thr Phe Phe Asn Leu Thr Ala Gly  
 165 170 175

Ser Lys Tyr Asn Ile Ala Ile Thr Ala Val Ser Gly Gly Lys Arg Ser  
 180 185 190

Phe Ser Val Tyr Thr Asn Gly Ser Thr Val Pro Ser Pro Val Lys Asp  
 195 200 205

Ile Gly Ile Ser Thr Lys Ala Asn Ser Leu Leu Ile Ser Trp Ser His  
 210 215 220





Gly Ser Gly Asn Val Glu Arg Tyr Arg Leu Met Leu Met Asp Lys Gly  
 225 230 235 240

Ile Leu Val His Gly Gly Val Val Asp Lys His Ala Thr Ser Tyr Ala  
 245 250 255

Phe His Gly Leu Ser Pro Gly Tyr Leu Tyr Asn Leu Thr Val Met Thr  
 260 265 270

Glu Ala Ala Gly Leu Gln Asn Tyr Arg Trp Lys Leu Val Arg Thr Ala  
 275 280 285

Pro Met Glu Val Ser Asn Leu Lys Val Thr Asn Asp Gly Ser Leu Thr  
 290 295 300

Ser Leu Lys Val Lys Trp Gln Arg Pro Pro Gly Asn Val Asp Ser Tyr  
 305 310 315 320

Asn Ile Thr Leu Ser His Lys Gly Thr Ile Lys Glu Ser Arg Val Leu  
 325 330 335

Ala Pro Trp Ile Thr Glu Thr His Phe Lys Glu Leu Val Pro Gly Arg  
 340 345 350

Leu Tyr Gln Val Thr Val Ser Cys Val Ser Gly Glu Leu Ser Ala Gln  
 355 360 365

Lys Met Ala Val Gly Arg Thr Phe Pro Asp Lys Val Ala Asn Leu Glu  
 370 375 380

Ala Asn Asn Asn Gly Arg Met Arg Ser Leu Val Val Ser Trp Ser Pro  
 385 390 395 400

Pro Ala Gly Asp Trp Glu Gln Tyr Arg Ile Leu Leu Phe Asn Asp Ser  
 405 410 415

Val Val Leu Leu Asn Ile Thr Val Gly Lys Glu Glu Thr Gln Tyr Val  
 420 425 430

Met Asp Asp Thr Gly Leu Val Pro Gly Arg Gln Tyr Glu Val Glu Val  
 435 440 445

Ile Val Glu Ser Gly Asn Leu Lys Asn Ser Glu Arg Cys Gln Gly Arg  
 450 455 460

Thr Val Pro Leu Ala Val Leu Gln Leu Arg Val Lys His Ala Asn Glu  
 465 470 475 480



Thr Ser Leu Ser Ile Met Trp Gln Thr Pro Val Ala Glu Trp Glu Lys  
 485 490 495

Tyr Ile Ile Ser Leu Ala Asp Arg Asp Leu Leu Leu Ile His Lys Ser  
 500 505 510

Leu Ser Lys Asp Ala Lys Glu Phe Thr Phe Thr Asp Leu Val Pro Gly  
 515 520 525

Arg Lys Tyr Met Ala Thr Val Thr Ser Ile Ser Gly Asp Leu Lys Asn  
 530 535 540

Ser Ser Ser Val Lys Gly Arg Thr Val Pro Ala Gln Val Thr Asp Leu  
 545 550 555 560

His Val Ala Asn Gln Gly Met Thr Ser Ser Leu Phe Thr Asn Trp Thr  
 565 570 575

Gln Ala Gln Gly Asp Val Glu Phe Tyr Gln Val Leu Leu Ile His Glu  
 580 585 590

Asn Val Val Ile Lys Asn Glu Ser Ile Ser Ser Glu Thr Ser Arg Tyr  
 595 600 605

Ser Phe His Ser Leu Lys Ser Gly Ser Leu Tyr Ser Val Val Val Thr  
 610 615 620

Thr Val Ser Gly Gly Ile Ser Ser Arg Gln Val Val Val Glu Gly Arg  
 625 630 635 640

Thr Val Pro Ser Ser Val Ser Gly Val Thr Val Asn Asn Ser Gly Arg  
 645 650 655

Asn Asp Tyr Leu Ser Val Ser Trp Leu Val Ala Pro Gly Asp Val Asp  
 660 665 670

Asn Tyr Glu Val Thr Leu Ser His Asp Gly Lys Val Val Gln Ser Leu  
 675 680 685

Val Ile Ala Lys Ser Val Arg Glu Cys Ser Phe Ser Ser Leu Thr Pro  
 690 695 700

Gly Arg Leu Tyr Thr Val Thr Ile Thr Thr Arg Ser Gly Lys Tyr Glu  
 705 710 715 720

Asn His Ser Phe Ser Gln Glu Arg Thr Val Pro Asp Lys Val Gln Gly  
 725 730 735



Val Ser Val Ser Asn Ser Ala Arg Ser Asp Tyr Leu Arg Val Ser Trp  
 740 745 750

Val Tyr Ala Thr Gly Asp Phe Asp His Tyr Glu Val Thr Ile Lys Asn  
 755 760 765

Lys Asn Asn Phe Ile Gln Thr Lys Ser Ile Pro Lys Ser Glu Asn Glu  
 770 775 780

Cys Val Phe Val Gln Leu Val Pro Gly Arg Leu Tyr Ser Val Thr Val  
 785 790 795 800

Thr Thr Lys Ser Gly Gln Tyr Glu Ala Asn Glu Gln Gly Asn Gly Arg  
 805 810 815

Thr Ile Pro Glu Pro Val Lys Asp Leu Thr Leu Arg Asn Arg Ser Thr  
 820 825 830

Glu Asp Leu His Val Thr Trp Ser Gly Ala Asn Gly Asp Val Asp Gln  
 835 840 845

Tyr Glu Ile Gln Leu Leu Phe Asn Asp Met Lys Val Phe Pro Pro Phe  
 850 855 860

His Leu Val Asn Thr Ala Thr Glu Tyr Arg Phe Thr Ser Leu Thr Pro  
 865 870 875 880

Gly Arg Gln Tyr Lys Ile Leu Val Leu Thr Ile Ser Gly Asp Val Gln  
 885 890 895

Gln Ser Ala Phe Ile Glu Gly Phe Thr Val Pro Ser Ala Val Lys Asn  
 900 905 910

Ile His Ile Ser Pro Asn Gly Ala Thr Asp Ser Leu Thr Val Asn Trp  
 915 920 925

Thr Pro Gly Gly Gly Asp Val Asp Ser Tyr Thr Val Ser Ala Phe Arg  
 930 935 940

His Ser Gln Lys Val Asp Ser Gln Thr Ile Pro Lys His Val Phe Glu  
 945 950 955 960

His Thr Phe His Arg Leu Glu Ala Gly Glu Gln Tyr Gln Ile Met Ile  
 965 970 975

Ala Ser Val Ser Gly Ser Leu Lys Asn Gln Ile Asn Val Val Gly Arg  
 980 985 990



Thr Val Pro Ala Ser Val Gln Gly Val Ile Ala Asp Asn Ala Tyr Ser  
 995 1000 1005

Ser Tyr Ser Leu Ile Val Ser Trp Gln Lys Ala Ala Gly Val Ala Glu  
 1010 1015 1020

Arg Tyr Asp Ile Leu Leu Leu Thr Glu Asn Gly Ile Leu Leu Arg Asn  
 1025 1030 1035 1040

Thr Ser Glu Pro Ala Thr Thr Lys Gln His Lys Phe Glu Asp Leu Thr  
 1045 1050 1055

Pro Gly Lys Lys Tyr Lys Ile Gln Ile Leu Thr Val Ser Gly Gly Leu  
 1060 1065 1070

Phe Ser Lys Glu Ala Gln Thr Glu Gly Arg Thr Val Pro Ala Ala Val  
 1075 1080 1085

Thr Asp Leu Arg Ile Thr Glu Asn Ser Thr Arg His Leu Ser Phe Arg  
 1090 1095 1100

Trp Thr Ala Ser Glu Gly Glu Leu Ser Trp Tyr Asn Ile Phe Leu Tyr  
 1105 1110 1115 1120

Asn Pro Asp Gly Asn Leu Gln Glu Arg Ala Gln Val Asp Pro Leu Val  
 1125 1130 1135

Gln Ser Phe Ser Phe Gln Asn Leu Leu Gln Gly Arg Met Tyr Lys Met  
 1140 1145 1150

Val Ile Val Thr His Ser Gly Glu Leu Ser Asn Glu Ser Phe Ile Phe  
 1155 1160 1165

Gly Arg Thr Val Pro Ala Ser Val Ser His Leu Arg Gly Ser Asn Arg  
 1170 1175 1180

Asn Thr Thr Asp Ser Leu Trp Phe Asn Trp Ser Pro Ala Ser Gly Asp  
 1185 1190 1195 1200

Phe Asp Phe Tyr Glu Leu Ile Leu Tyr Asn Pro Asn Gly Thr Lys Lys  
 1205 1210 1215

Glu Asn Trp Lys Asp Lys Asp Leu Thr Glu Trp Arg Phe Gln Gly Leu  
 1220 1225 1230

Val Pro Gly Arg Lys Tyr Val Leu Trp Val Val Thr His Ser Gly Asp  
 1235 1240 1245





Leu Ser Asn Lys Val Thr Ala Glu Ser Arg Thr Ala Pro Ser Pro Pro  
 1250 1255 1260

Ser Leu Met Ser Phe Ala Asp Ile Ala Asn Thr Ser Leu Ala Ile Thr  
 1265 1270 1275 1280

Trp Lys Gly Pro Pro Asp Trp Thr Asp Tyr Asn Asp Phe Glu Leu Gln  
 1285 1290 1295

Trp Leu Pro Arg Asp Ala Leu Thr Val Phe Asn Pro Tyr Asn Asn Arg  
 1300 1305 1310

Lys Ser Glu Gly Arg Ile Val Tyr Gly Leu Arg Pro Gly Arg Ser Tyr  
 1315 1320 1325

Gln Phe Asn Val Lys Thr Val Ser Gly Asp Ser Trp Lys Thr Tyr Ser  
 1330 1335 1340

Lys Pro Ile Phe Gly Ser Val Arg Thr Lys Pro Asp Lys Ile Gln Asn  
 1345 1350 1355 1360

Leu His Cys Arg Pro Gln Asn Ser Thr Ala Ile Ala Cys Ser Trp Ile  
 1365 1370 1375

Pro Pro Asp Ser Asp Phe Asp Gly Tyr Ser Ile Glu Cys Arg Lys Met  
 1380 1385 1390

Asp Thr Gln Glu Val Glu Phe Ser Arg Lys Leu Glu Lys Glu Lys Ser  
 1395 1400 1405

Leu Leu Asn Ile Met Met Leu Val Pro His Lys Arg Tyr Leu Val Ser  
 1410 1415 1420

Ile Lys Val Gln Ser Ala Gly Met Thr Ser Glu Val Val Glu Asp Ser  
 1425 1430 1435 1440

Thr Ile Thr Met Ile Asp Arg Pro Pro Pro Pro Pro Pro His Ile Arg  
 1445 1450 1455

Val Asn Glu Lys Asp Val Leu Ile Ser Lys Ser Ser Ile Asn Phe Thr  
 1460 1465 1470

Val Asn Cys Ser Trp Phe Ser Asp Thr Asn Gly Ala Val Lys Tyr Phe  
 1475 1480 1485

Thr Val Val Val Arg Glu Ala Asp Gly Ser Asp Glu Leu Lys Pro Glu  
 1490 1495 1500



Gln Gln His Pro Leu Pro Ser Tyr Leu Glu Tyr Arg His Asn Ala Ser			
1505	1510	1515	1520
Ile Arg Val Tyr Gln Thr Asn Tyr Phe Ala Ser Lys Cys Ala Glu Asn			
	1525	1530	1535
Pro Asn Ser Asn Ser Lys Ser Phe Asn Ile Lys Leu Gly Ala Glu Met			
	1540	1545	1550
Glu Ser Leu Gly Gly Lys Cys Asp Pro Thr Gln Gln Lys Phe Cys Asp			
	1555	1560	1565
Gly Pro Leu Lys Pro His Thr Ala Tyr Arg Ile Ser Ile Arg Ala Phe			
	1570	1575	1580
Thr Gln Leu Phe Asp Glu Asp Leu Lys Glu Phe Thr Lys Pro Leu Tyr			
	1585	1590	1595
Ser Asp Thr Phe Phe Ser Leu Pro Ile Thr Thr Glu Ser Glu Pro Leu			
	1605	1610	1615
Phe Gly Ala Ile Glu Gly Val Ser Ala Gly Leu Phe Leu Ile Gly Met			
	1620	1625	1630
Leu Val Ala Val Val Ala Leu Leu Ile Cys Arg Gln Lys Val Ser His			
	1635	1640	1645
Gly Arg Glu Arg Pro Ser Ala Arg Leu Ser Ile Arg Arg Asp Arg Pro			
	1650	1655	1660
Leu Ser Val His Leu Asn Leu Gly Gln Lys Gly Asn Arg Lys Thr Ser			
	1665	1670	1675
Cys Pro Ile Lys Ile Asn Gln Phe Glu Gly His Phe Met Lys Leu Gln			
	1685	1690	1695
Ala Asp Ser Asn Tyr Leu Leu Ser Lys Glu Tyr Glu Glu Leu Lys Asp			
	1700	1705	1710
Val Gly Arg Asn Gln Ser Cys Asp Ile Ala Leu Leu Pro Glu Asn Arg			
	1715	1720	1725
Gly Lys Asn Arg Tyr Asn Asn Ile Leu Pro Tyr Asp Ala Thr Arg Val			
	1730	1735	1740
Lys Leu Ser Asn Val Asp Asp Asp Pro Cys Ser Asp Tyr Ile Asn Ala			
	1745	1750	1755
			1760



Ser Tyr Ile Pro Gly Asn Asn Phe Arg Arg Glu Tyr Ile Val Thr Gln  
 1765 1770 1775  
 Gly Pro Leu Pro Gly Thr Lys Asp Asp Phe Trp Lys Met Val Trp Glu  
 1780 1785 1790  
 Gln Asn Val His Asn Ile Val Met Val Thr Gln Cys Val Glu Lys Gly  
 1795 1800 1805  
 Arg Val Lys Cys Asp His Tyr Trp Pro Ala Asp Gln Asp Ser Leu Tyr  
 1810 1815 1820  
 Tyr Gly Asp Leu Ile Leu Gln Met Leu Ser Glu Ser Val Leu Pro Glu  
 1825 1830 1835 1840  
 Trp Thr Ile Arg Glu Phe Lys Ile Cys Gly Glu Glu Gln Leu Asp Ala  
 1845 1850 1855  
 His Arg Leu Ile Arg His Phe His Tyr Thr Val Trp Pro Asp His Gly  
 1860 1865 1870  
 Val Pro Glu Thr Thr Gln Ser Leu Ile Gln Phe Val Arg Thr Val Arg  
 1875 1880 1885  
 Asp Tyr Ile Asn Arg Ser Pro Gly Ala Gly Pro Thr Val Val His Cys  
 1890 1895 1900  
 Ser Ala Gly Val Gly Arg Thr Gly Thr Phe Ile Ala Leu Asp Arg Ile  
 1905 1910 1915 1920  
 Leu Gln Gln Leu Asp Ser Lys Asp Ser Val Asp Ile Tyr Gly Ala Val  
 1925 1930 1935  
 His Asp Leu Arg Leu His Arg Val His Met Val Gln Thr Glu Cys Gln  
 1940 1945 1950  
 Tyr Val Tyr Leu His Gln Cys Val Arg Asp Val Leu Arg Ala Arg Lys  
 1955 1960 1965  
 Leu Arg Ser Glu Gln Glu Asn Pro Leu Phe Pro Ile Tyr Glu Asn Val  
 1970 1975 1980  
 Asn Pro Glu Tyr His Arg Asp Pro Val Tyr Ser Arg His  
 1985 1990 1995

&lt;210&gt; 55

&lt;211&gt; 453



&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 55

Met Lys Leu Leu Val Ile Leu Leu Phe Ser Gly Leu Ile Thr Gly Phe  
 1 5 10 15

Arg Ser Asp Ser Ser Ser Ser Leu Pro Pro Lys Leu Leu Leu Val Ser  
 20 25 30

Phe Asp Gly Phe Arg Ala Asp Tyr Leu Lys Asn Tyr Glu Phe Pro His  
 35 40 45

Leu Gln Asn Phe Ile Lys Glu Gly Val Leu Val Glu His Val Lys Asn  
 50 55 60

Val Phe Ile Thr Lys Thr Phe Pro Asn His Tyr Ser Ile Val Thr Gly  
 65 70 75 80

Leu Tyr Glu Glu Ser His Gly Ile Val Ala Asn Ser Met Tyr Asp Ala  
 85 90 95

Val Thr Lys Lys His Phe Ser Asp Ser Asn Asp Lys Asp Pro Phe Trp  
 100 105 110

Trp Asn Glu Ala Val Pro Ile Trp Val Thr Asn Gln Leu Gln Glu Asn  
 115 120 125

Arg Ser Ser Ala Ala Ala Met Trp Pro Gly Thr Asp Val Pro Ile His  
 130 135 140

Asp Thr Ile Ser Ser Tyr Phe Met Asn Tyr Asn Ser Ser Val Ser Phe  
 145 150 155 160

Glu Glu Arg Leu Asn Asn Ile Thr Met Trp Leu Asn Asn Ser Asn Pro  
 165 170 175

Pro Val Thr Phe Ala Thr Leu Tyr Trp Glu Glu Pro Asp Ala Ser Gly  
 180 185 190

His Lys Tyr Gly Pro Glu Asp Lys Glu Asn Met Ser Arg Val Leu Lys  
 195 200 205

Lys Ile Asp Asp Leu Ile Gly Asp Leu Val Gln Arg Leu Lys Met Leu  
 210 215 220

Gly Leu Trp Glu Asn Leu Asn Val Ile Ile Thr Ser Asp His Gly Met  
 225 230 235 240





Thr Gln Cys Ser Gln Asp Arg Leu Ile Asn Leu Asp Ser Cys Ile Asp  
 245 250 255  
 His Ser Tyr Tyr Thr Leu Ile Asp Leu Ser Pro Val Ala Ala Ile Leu  
 260 265 270  
 Pro Lys Ile Asn Arg Thr Glu Val Tyr Asn Lys Leu Lys Asn Cys Ser  
 275 280 285  
 Pro His Met Asn Val Tyr Leu Lys Glu Asp Ile Pro Asn Arg Phe Tyr  
 290 295 300  
 Tyr Gln His Asn Asp Arg Ile Gln Pro Ile Ile Leu Val Ala Asp Glu  
 305 310 315 320  
 Gly Trp Thr Ile Val Leu Asn Glu Ser Ser Gln Lys Leu Gly Asp His  
 325 330 335  
 Gly Tyr Asp Asn Ser Leu Pro Ser Met His Pro Phe Leu Ala Ala His  
 340 345 350  
 Gly Pro Ala Phe His Lys Gly Tyr Lys His Ser Thr Ile Asn Ile Val  
 355 360 365  
 Asp Ile Tyr Pro Met Met Cys His Ile Leu Gly Leu Lys Pro His Pro  
 370 375 380  
 Asn Asn Gly Thr Phe Gly His Thr Lys Cys Leu Leu Val Asp Gln Trp  
 385 390 395 400  
 Cys Ile Asn Leu Pro Glu Ala Ile Ala Ile Val Ile Gly Ser Leu Leu  
 405 410 415  
 Val Leu Thr Met Leu Thr Cys Leu Ile Ile Ile Met Gln Asn Arg Leu  
 420 425 430  
 Ser Val Pro Arg Pro Phe Ser Arg Leu Gln Leu Gln Glu Asp Asp Asp  
 435 440 445  
 Asp Pro Leu Ile Gly  
 450

&lt;210&gt; 56

&lt;211&gt; 537

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens



&lt;400&gt; 56

Met Ser Lys Pro His Ser Glu Ala Gly Thr Ala Phe Ile Gln Thr Gln  
 1 5 10 15

Gln Leu His Ala Ala Met Ala Asp Thr Phe Leu Glu His Met Cys Arg  
 20 25 30

Leu Asp Ile Asp Ser Pro Pro Ile Thr Ala Arg Asn Thr Gly Ile Ile  
 35 40 45

Cys Thr Ile Gly Pro Ala Ser Arg Ser Val Glu Thr Leu Lys Glu Met  
 50 55 60

Ile Lys Ser Gly Met Asn Val Ala Arg Leu Asn Phe Ser His Gly Thr  
 65 70 75 80

His Glu Tyr His Ala Glu Thr Ile Lys Asn Val Arg Thr Ala Thr Glu  
 85 90 95

Ser Phe Ala Ser Asp Pro Ile Leu Tyr Arg Pro Val Ala Val Ala Leu  
 100 105 110

Asp Thr Lys Gly Pro Glu Ile Arg Thr Gly Leu Ile Lys Gly Ser Gly  
 115 120 125

Thr Ala Glu Val Glu Leu Lys Lys Gly Ala Thr Leu Lys Ile Thr Leu  
 130 135 140

Asp Asn Ala Tyr Met Glu Lys Cys Asp Glu Asn Ile Leu Trp Leu Asp  
 145 150 155 160

Tyr Lys Asn Ile Cys Lys Val Val Glu Val Gly Ser Lys Ile Tyr Val  
 165 170 175

Asp Asp Gly Leu Ile Ser Leu Gln Val Lys Gln Lys Gly Ala Asp Phe  
 180 185 190

Leu Val Thr Glu Val Glu Asn Gly Gly Ser Leu Gly Ser Lys Lys Gly  
 195 200 205

Val Asn Leu Pro Gly Ala Ala Val Asp Leu Pro Ala Val Ser Glu Lys  
 210 215 220

Asp Ile Pro Gly Ser Glu Ser Leu Gly Val Glu Gln Asp Val Asp Met  
 225 230 235 240

Val Phe Ala Ser Phe His Pro Ala Lys Ala Ser Gly Cys Pro Met Glu



245	250	255
Ala Leu Gly Ala Val Leu Gly Arg Glu Gly Lys Arg Asn Ile Lys Ile		
260	265	270
Ile Ser Lys Ile Glu Asn His Glu Gly Val Arg Arg Phe Asp Glu Ile		
275	280	285
Leu Glu Ala Ser Asp Gly Ile Met Val Ala Arg Gly Asp Leu Gly Ile		
290	295	300
Glu Ile Pro Ala Glu Lys Val Phe Leu Ala Gln Lys Met Met Ile Gly		
305	310	315
Arg Cys Asn Pro Arg Thr Gly Lys Pro Val Ile Cys Ala Thr Gln Met		
325	330	335
Leu Glu Ser Ile Ile Lys Lys Pro Arg Pro Thr Arg Ala Glu Gly Ser		
340	345	350
Asp Val Ala Asn Ala Val Leu Asp Gly Ala Asp Cys Ile Met Leu Ser		
355	360	365
Gly Glu Thr Ala Lys Gly Asp Tyr Pro Leu Glu Ala Val Arg Met Gln		
370	375	380
His Leu Ile Ala Arg Glu Ala Glu Ala Ala Ile Tyr His Leu Gln Leu		
385	390	395
Phe Glu Glu Leu Arg Arg Leu Ala Pro Ile Thr Ser Asp Pro Thr Glu		
405	410	415
Ala Thr Ala Val Gly Ala Val Glu Ala Ser Phe Lys Cys Cys Ser Gly		
420	425	430
Ala Ile Ile Val Leu Thr Lys Ser Gly Arg Ser Ala His Gln Val Ala		
435	440	445
Arg Tyr Arg Pro Arg Ala Pro Ile Ile Ala Val Thr Arg Asn Pro Gln		
450	455	460
Thr Ala Arg Gln Ala His Leu Tyr Arg Gly Ile Phe Pro Val Leu Cys		
465	470	475
Lys Asp Pro Val Gln Glu Ala Trp Ala Glu Asp Val Asp Leu Arg Val		
485	490	495
Asn Phe Ala Met Asn Val Gly Lys Ala Arg Gly Phe Phe Lys Lys Gly		



500

505

510

Asp Val Val Ile Val Leu Thr Gly Trp Arg Pro Gly Ser Gly Phe Thr  
 515 520 525

Asn Thr Met Arg Val Val Pro Val Pro  
 530 535

